PATENT
Attorney Docket No. 10142.0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	
Donald GULLBERG	Group Art Unit: 1644
Application No.: 09/980,403	Examiner: Haddad, Maher M.
Filed: April 15, 2002	
For: AN INTEGRIN HETERODIMER AND AN ALPHA SUBUNIT THEREOF	Confirmation No.: 3147

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

DECLARATION UNDER 37 C.F.R. § 1.132

I, Staffan Johansson, do hereby make the following declaration:

1. I am a Professor of Medical Cell Biology at the University of Uppsala.

2. My *curriculum vitae* is provided as Appendix 1.

3. I have no personal interest in the outcome of the prosecution of US Patent Application No. 09/980,403.

4. I have actively worked in the field of integrin research for the last 20 years.

5. I have been asked to comment on the following publications with respect to their reception in the field of integrin research at the time of their publication and their relevance to the field:

Gullberg et al., *Dev. Dyn.* 204:57-65 (1995)

Velling et al., *J. Biol. Chem.* 274:25735-42 (1999)

I have read and understood these publications.

6. The field of integrin research, and specifically integrin gene cloning, was one of the most competitive fields in biomedical research in the 1990's. The integrins were found to be fundamental for the function of all cells in the body, as shown by the range of defects resulting from mutations or deletions in the different integrin genes. These defects include bleeding disorders, inflammatory disorders, cancer metastasis, lack of immune response, skin detachment (epidermis pilosa), and muscle dystrophy[1]. Therefore, the identification and cloning of each integrin subunit was considered a major event in biomedical research, which in many cases opened up a whole new area of investigation.

7. New integrin subunits were identified and cloned at a rapid pace in the early 1990's, mostly by techniques that exploited the regions of homology between already known integrin subunits, but by 1995 the discovery of new integrin subunits had slowed dramatically. My assessment in 1995 and the subsequent years was that possibly all integrin subunits had been identified and cloned. This was also a general view in the integrin research community[2]. This belief was based on the sharp drop in the rate of discovery of new integrin subunits in spite of improved knowledge, reagents, and methods, such as advanced knowledge of conserved sequences and domains of

---

1 See, e.g., Hynes, R.O., Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992 Apr 3;69(1):11-25. All references cited in this declaration are provided as Appendix 2.

integrins, the availability of PCR technology and integrin specific antibodies, and improved knowledge of the conditions for co-immunoprecipitation and affinity chromatography of integrins. In addition, reports -- either in publications[3], scientific meetings or informal discussions with colleagues -- indicating the presence of additional, yet unidentified integrins had ceased.

8. The only exception I knew of at that time was Donald Gullberg and his publication of certain bands in polyacrylamide gels, which he felt may represent an integrin a subunit which did not correlate to any of the known integrins. Based on these gel bands Gullberg et al postulated the existence of a new integrin a subunit which they termed integrin αmt. See Gullberg et al., *Dev. Dyn.* 204:57-65 (1995).

9. Gullberg's report was met with skepticism in the integrin research community, including myself. As explained in 7., the general belief was that possibly all integrins had been identified and cloned at that point. Gullberg's proposal of an additional, yet unidentified integrin α chain based solely on biochemical experiments, but without providing any sequence of the postulated molecule, seemed to be scientifically weak. The complexity of the integrin family was well recognized at that time[4], and Gullberg simply did not provide any evidence that would have distinguished his gel bands from known integrins or from alternative splicing or glycosylation products of known integrins. Actually, Gullberg did not provide any hard evidence that his gel

---

2 See, e.g., Hynes, R.O., Cell adhesion: old and new questions. Trends Cell Biol. 1999 Dec;9(12):M33-7.

3 See, e.g., Hynes, R.O., Cell adhesion: old and new questions. Trends Cell Biol. 1999 Dec;9(12):M33-7.

bands correlated to any integrin at all. There was only circumstantial evidence in the molecular size and the co-immunoprecipitation with integrin β1.

10. Since the discovery of the first integrins, odd reports of "integrin-like" genes or proteins from various sources had appeared regularly. Such reports related to "integrins" in plants and yeast, single chain "integrins", and truncated forms of integrins[5]. In many cases these reports turned out not to describe true integrins. For example, it is now known that integrins do not exist in plants and yeast[6]. There was a general suspicion in the research community towards these odd reports and in the absence of convincing evidence the experts in the field had a healthy critical view towards any claims of a new integrin in the absence of sequence data. I have experienced this general scepticism myself when my initial attempts to publish biochemical data on the isolation of a new integrin α chain were rejected by the peer reviewers of two different journals. In the absence of complete sequence data my peers simply were unconvinced that we had actually identified a novel integrin α subunit[7]. Only after obtaining more sequence information we were able to publish the characterization of the novel integrin subunit

---

4 See, e.g., Hynes, R.O., Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992 Apr 3;69(1):11-25; Loftus, J.C., et al., Integrin-mediated cell adhesion: the extracellular face. J Biol Chem. 1994 Oct 14;269(41):25235-8.

5 See, e.g., Gale, C., et al., Cloning and expression of a gene encoding an integrin-like protein in Candida albicans. Proc Natl Acad Sci USA. 1996 Jan 9;93(1):357-61; Berg, R.W., et al., Cloning and characterization of a novel beta integrin-related cDNA coding for the protein TIED ("ten beta integrin EGF-like repeat domains") that maps to chromosome band 13q33: A divergent stand-alone integrin stalk structure. Genomics. 1999 Mar 1;56(2):169-78; Laval, V., et al., A family of Arabidopsis plasma membrane receptors presenting animal beta-integrin domains. Biochim Biophys Acta. 1999 Nov 16;1435(1-2):61-70.

6 See, e.g., Rubin, G.M., et al., Comparative genomics of the eukaryotes. Science. 2000 Mar 24;287(5461):2204-15.

7 An example of a relevant communication by a peer reviewer is provided as Appendix 3.

α9[8]. This general suspicion and scepticism also applied to Gullberg's report of integrin amt, even though he somehow managed to get his biochemical data published.

11. The obvious way to clone Gullberg's postulated novel integrin for one of ordinary skill in the art was to take G6 muscle cells as a source material for construction of a library (cDNA or expression library), for isolation of mRNA for PCR amplification, or for affinity purification of the target protein for amino terminal sequencing. One could also use the cells to raise antibodies against cell surface molecules, including integrins, and select antibodies based on their ability to block adhesion of G6 cells to extracellular matrix proteins. Such antibodies could then be used for affinity purification of the target protein or for screening of an expression library. Probes for cDNA library screening and primers for PCR amplification would have been designed based on the homology regions of known integrin subunits. However, to my best knowledge no one, including Gullberg himself, succeeded in cloning the postulated new integrin amt by any of these methods. After having failed to clone the postulated integrin amt by any of these methods one of ordinary skill in the art would have concluded that most likely there was no novel integrin present in these muscle cells. This person would have had no good reason to experiment randomly with other source materials since no evidence for the expression of the postulated integrin amt in other cell types existed and the expectation of success would have been dramatically reduced because of the known cell type-, differentiation state-, and developmental stage-specific expression and regulation

---

8 Forsberg, E., et al., Purification and characterization of integrin alpha 9 beta 1. Exp Cell Res. 1994 Jul;213(1):183-90.

patterns of integrins[9]. Thus, there was no good reason to try to clone integrin αmt from uterus tissue.

12. The Velling et al report in 1999 of the cloning of the new integrin α11 and its correlation to the previously reported integrin αmt came as a surprise to the field, for all the reasons outlined above. However, despite the statement by Velling et al that the newly cloned integrin α11 correlates to the integrin αmt of Gullberg et al this has never been clearly demonstrated. The correlation is based purely on a few biophysical properties (see Velling et al.) but not on any sequence identity. Gullberg's integrin αmt has to the best of my knowledge never been cloned from myotubes, thereby precluding any sequence comparison. Hence, whether integrin αmt identified by Gullberg et al is identical to integrin α11 identified by Velling et al remains unknown to this day.

13. As explained in 6., the identification and cloning of each integrin subunit was considered a major event in biomedical research, which in many cases opened a whole new area of investigation after identification of each unique integrin function. Examples are αIIβ3 (expressed on thrombocytes, important regulatory role in blood coagulation), αLβ2 (role in the immune system, deficiency gives LAD, leucocyte adhesion deficiency), α6β4 (expressed on hemidesmosomes, deficiency gives epidermolysis bullosa), αVβ6 (regulates TGFbeta, important regulatory role for the immune system and matrix turnover[10]).

---

9 See, e.g., Hynes, R.O., Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992 Apr 3;69(1):1 1-25.

10 See, e.g., Munger, J.S., et al., The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell. 1999 Feb 5;96(3):319-28.

14. I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated: January 16, 2008 By: Steffan Johansson

APPENDIX 1

OF
DECLARATION UNDER 37 C.F.R. § 1.132

DATED JANUARY 16, 2008

BY
STAFFAN JOHANSSON

Content:

Curriculum Vitae of Staffan Johansson (10 pages)

CURRICULUM VITAE

Staffan Johansson, 530221-1031

Address Nordhemsvägen 8B, 756 46 Uppsala
Tel. 018-302734 (home)

Family Lena Kjellen (co-habitant)
Alva Kjellén (910420)
Erland Kjellén (940604)

Affiliation Dept. of Medical Biochemistry and Microbiology, Uppsala Univ.

Examina BS (FK) 1977 at the University of Uppsala: Chemistry 40p, Biology 100p

PhD 830509 with the thesis "Pericellular matrix components and cell adhesion" at the Dept. of Med. and Physiol. Chem., Univ. of Uppsala

Docent 880525 in medical and physiological chemistry, Univ. of Uppsala.

Positions Professor in medical cell biology, Univ. of Uppsala, 000101-.

Assoc. professor in medical cell biology, Univ. of Uppsala, 990302-991231.

Assoc. professor in the field of "Zoological cell biology", defrayed by the National Science Research Council, at the Dept. of Med. and Physiol. Chem., Univ. of Uppsala, 921201-990228.

Leave of absence (taking care of children), 50% of 11 months during the period 941121-951231.

Various positions at the Dept. of Med. and Physiol. Chem., Univ. of Uppsala, 830901-921130.

Visiting research associate at the Connective Tissue Laboratory, University of Alabama in Birmingham, USA, 791203-810518.

PhD position, 780701-830831.

Other assignments

Principal PhD supervisor for:
Erik Forsberg (PhD exam 931203),
Peter McCourt (PhD exam 990226)
Krister Wennerberg (PhD exam 990416)
Gunbjørg Svineng (PhD exam 991216)
Lars Lohikangas, (PhD exam 000530)
Annika Armulik (PhD exam 001103)

Stina Nilsson (PhD exam 060505)
Anne Simonsson (PhD exam 2007-10-19)

Associated Postdoctoral fellows:
Tom Prasthofer (1989-1992)
Nancy Martin (1989-1992)
Paola Longati (2001/2002)
Jian-He Wu (2002/2003)
Sophie Johansson 2001-2003
Teet Velling (2002-2004)
Nathalie Bot (2006/2007

Thesis opponent for:
Tore Gauperaa, Univ. of Tromsø, Norway, 860524
Yngve Sommarin, Univ. of Lund, 870123
Lars Uhlin-Hansen, Univ. of Tromsø, Norway, 900324
Ylva Härdig, MAS, Univ. of Lund, 961109
Lisbet Camper, Univ. of Lund, 000324
Christina Wenglén, Univ. of Lund, 011013
Ramin Massoumi, MAS, Univ. of Lund, 020524
Hao Yu, Univ. of Lund, 040602
Xiaojie Xian, Unif. of Göteborg. 50301
Lars Bryngelsson Ohlsson, Univ. of Lund, 051111
Helena Stensman, Univ. of Lund, 070608

Reviewer of articles for:
J. Cell Biol., Mol. Cell Biol., J. Biol. Chem., Oncogene, Exp. Cell Res., J. Cell. Biochem., Eur. J. Biochem., J. Clin. Invest., Matrix Biol., Thromb. Res., FEBS Lett.

Evaluator of grant applications for
Cancerfonden, Sweden, 2007-
Swedich Medical Research Council, 2000-2007
The Alzheimer´s Association, 1999-
EU, Cancer, 2003
The Wellcome Trust, UK, 2003
Human Frontier Science Program, 1999.
The Cell Biol. Research Program of the Academy of Finland, 1998.

Evaluator of positions
Docentur, 5 cases
Lectorship, Södertörn Högskola, 2001
Lectorships (two), Lunds Univ. 2007
Senior VR position, Mol. Cell Biol. 2007

Present major external grants

Swedich Research Council (Medicin), 1985-

Swedish Cancer Foundation, 1998-

PUBLICATIONS

Original articles

1. Rubin, K., **Johansson, S.**, Pettersson, I., Ocklind, C., Öbrink, B., Höök, M. 1979. "Attachment of rat hepatocytes to collagen and fibronectin: a study using antibodies against the surface components" *Biochem. Biophys. Commun.* 91, 86-94

2. **Johansson, S.**, Rubin, K., Höök, M., Ahlgren T., Seljelid, R. 1979. "In vitro biosynthesis of cold insoluble globulin (fibronectin)". *FEBS Lett.* 105, 313-316

3. Hedman, K., Kurkinen, M., Alitalo, K., Vaheri, A., **Johansson, S.**, Höök, M. 1979. "Isolation of the pericellular matrix of human fibroblast cultures". *J. Cell Biol.* 81, 83-91

4. **Johansson, S.**, Höök, M. 1980. "Heparin enhances the rate of binding of fibronectin to collagen". *Biochem. J.* 187, 521-524

5. Rubin, K., **Johansson, S.**, Höök, M., Öbrink, B. 1981. "Substrate adhesion of rat hepatocytes: On the role of fibronectin in cell spreading". *Exp. Cell Res.* 135, 127-135

6. **Johansson, S.**, Kjellén, L., Höök, M., Timple, R. 1981. "Substrate adhesion of rat hepatocytes: A comparison of laminin and fibronectin as attachment proteins". *J. Cell Biol.* 90, 260-264

7. Hedman, K., **Johansson, S.**, Vartio, T., Kjellén, L., Vaheri, A., Höök, M. 1982. "Structure of the pericellular matrix in human fibroblast cultures: Association of heparan and chondroitin sulfates with the fibronectin-procollagen fibers". *Cell* 26, 663-671

8. Linde, A., **Johansson, S.**, Jonsson, R., Jontell, M., 1982. "Localization of fibronectin during dentogenesis in rat incisor". *Arch. Oral. Biol.* 27, 1069-1073

9. Bagge, L., Hedstrand, U., Höök, M., **Johansson, S.**, Lind, E., Modig, J., Saldeen, T. 1983. "Fibrinolysis inhibition and fibronectin in the blood in patients with delayed microembolism syndrome". *Upsala J. Med. Sci.* 88, 81-94

10. Timpl, R., **Johansson, S.**, van Delden, V., Oberäumer, I., Höök, M. 1983. "Characterization of protease-resistant fragments of laminin mediating attachment and spreading of rat hepatocytes". *J. Biol. Chem.* 258, 8922-8927

11. **Johansson, S.** 1983. "Pericellular matrix components and cell adhesion". *Doctoral thesis at the University of Uppsala Sweden.*

12. **Johansson, S.**, Höök, M. 1984. "Substrate adhesion of rat hepatocytes: On the mechanism of cell attachment to fibronectin". *J. Cell Biol.* 98, 810-817

13. Hedman, K., Vartio, T., **Johansson, S.**, Kjellén, L., Höök, M., Linker, A., Salonen, E-M., Vaheri, A. 1984. "Integrity of the pericellular fibronectin matrix of fibroblasts is independent of sulfated glycosaminoglycans". *EMBO J.* 3, 581-584

14. **Johansson, S.**, Hedman, K., Kjellén, L., Christner, J., Vaheri, A., Höök M. 1985. "Structure and interactions of proteoglycans in the extracellular matrix produced by cultured human fibroblasts". *Biochem. J.* 232, 161-168

15. Smedsröd, B., **Johansson, S.**, Pertoft, H. 1985. "In vivo and in vitro studies on the uptake and degradation of soluble collagen in rat liver endothelial and Kupffer cells". *Biochem. J.* 228, 415-424

16. **Johansson, S.** 1985. "Demonstration of high affininty fibronectin-receptors on rat hepatocytes in suspension". *J. Biol. Chem.* 260, 1557-1561

17. **Johansson, S.**, Smedsröd, B. 1986 "Identification of a 72 kDa plasma gelatinase in preparations of fibronectin". *J. Biol. Chem.* 261, 4363-4366

18. Woods, A., Couchman, J., **Johansson, S.**, Höök, M. 1986. "Adhesion and cytoskeletal organization of fibroblasts in response to fibronectin fragments". *EMBO J.* 5, 65-670

19. **Johansson, S.**, Forsberg, E., Lundgren, B. 1987. "Comparison of fibronectin receptors from rat hepatocytes and fibroblasts". *J. Biol. Chem.* 262, 7819-7824

20. Perris, R., **Johansson, S.** 1987. "Amphibian neural crest cell migration on purified extracellular matrix components: A chondroitin sulfate proteoglycan inhibits locomotion on fibronectin substrates". *J. Cell Biol.* 105, 2511-2521

21. **Johansson, S.**, Gustafsson, S., Pertoft, H. 1987. "Identification of a fibronectin receptor specific for liver endothelial cells". *Exp. Cell Res.* 425-431

22. Wiersma, E. J., Fröman, G., **Johansson, S.**, Wadström, T. 1987. "Carbohydrate specific binding of fibronectin to Vibrio Cholerae cells". *FEMS Lett.* 44, 365-369

23. LeBaron, R., Esko, J., Woods, A., **Johansson, S.**, Höök, M. 1988. "Adhesion of glycosaminoglycan-deficient Chinese Hamster Ovary cell mutants to fibronectin substrate". *J. Cell Biol.* 106, 945-952

24. Woods, A., **Johansson, S.**, Höök, M. 1988. "Fibronectin fibril formation involves reorganization of external fibronectin by two cell surface components". *Exp. Cell Res.* 17, 272-283

25. Carri, N. G., Perris, R., **Johansson, S.**, Ebendal, T. 1988. "Differential promotion of neurite outgrowth from retinal explants by purifies extracellular matrix molecules". *J. Neurosci. Res.* 19, 428-439

26. Hedin, U., Bottger, B. A., Forsberg, E., **Johansson, S.**, Thyberg, J. 1988. "Diverse effects of fibronectin and laminin on phenotypic properties of cultured arterial smooth muscle cells. *J. Cell Biol.* 107, 307-319

27. Hedin, U., Bottger, B. A., Luthman, J., **Johansson, S.**, Thyberg, J. 1989. "A substrate of the cell-attachment sequence of fibronectin (Arg-Gly-Asp) is sufficient to promote the transition of arterial smooth muscle cells from a contractile to a synthetic phenotype". *Dev. Biol.* 133, 489-501

28. Bottger, B. A., Hedin, U., **Johansson, S.**, Thyberg, J. 1989. "Integrin-type fibronectin receptors of rat arterial smooth muscle cells: isolation, partial characterization and role in cytoskeletal organization and control of differentiated properties". *Differentiation* 41, 158-167

29. Smedsrød, B., Paulsson, M., **Johansson, S.** 1989. "Uptake and degradation in vivo and in vitro of laminin and nidogen by rat liver cells". *Biochem. J.* 261, 37-42

30. Perris, R., **Johansson, S.** 1990. "Inhibition of neural crest cell migration by aggregating chondroitin sulfate proteoglycans is mediated by their hyaluronan binding region". *Dev. Biol.* 137, 1-12

31. Hansson, M., Odin, P., **Johansson, S.**, Öbrink, B. 1990. "Comparison and functional characterization of C-CAM, glycoprotein IIb/IIIa and integrin beta-1 in rat platelets". *Thromb. Res.* 58, 61-73

32. Forsberg, E., Paulsson, M., Timpl, R., **Johansson, S.** 1990. "Characterization of a laminin receptor on rat hepatocytes". *J. Biol. Chem.* 265, 6376-6381

33. Stamatoglou, S. C., Sullivan, K. H., **Johansson, S.**, Bayley, P. M., Burdett, I. D., Hughes. 1990 "Localization of two fibronectin-binding glycoproteins in rat liver and primary hepatocytes". *J. Cell Sci.* 97, 595-606

34. Stamatoglou, S. C., Bawumia, S., **Johansson, S.**, Forsberg, E., Hughes, C. R. 1991. "Affinity of integrin $\alpha_1\beta_1$ from liver sinusoidal membranes for type IV collagen". *FEBS Lett.* 288, 241-243

35. Wärmegård, B., Martin, N., **Johansson** S. 1992. "cDNA-cloning and sequencing of rat alfa-1-macroglobulin". *Biochemistry* 31, 2346-2352

36. Pujades, C., Forsberg, E., Enrich, C., **Johansson, S.** 1992. "Changes in cell surface expression of fibronectin and fibronectin receptor during liver regeneration". *J. Cell Sci.* 102, 815-820

37. Smilenov, L., Forsberg, E., Zeligman, I., Sparrman, M., **Johansson, S.** 1992. "Separation of fibronectin from a plasma gelatinase using immobilized metal affinity chromatography". *FEBS Lett.* 302, 227-230

38. Murtomäki, S., Risteli, J., Risteli, L., Koivisto, U.-M., **Johansson, S.**, Liesi, P. 1992, "Laminin and its neurite outgrowth promoting domain accumulate in the brain in Alzheimer´s disease and Down´s syndrome". *J. Neurosci. Res.* 32, 261-273

39. Malmström, P.-U., Larsson, A., **Johansson, S.** 1993. "Urinary fibronectin in diagnosis and follow-up of patients with urinary bladder cancer". *Brittish J. Urol.* 72, 307-310

40. Forsberg, E., Ek, B., **Johansson, S.** 1994. "Purification and characterization of integrin $\alpha_9\beta_1$". *Exp. Cell Res.* 213, 183-190

41. Forsberg, E., Lindblom, A., Paulsson, M., **Johansson, S.** 1994. "Laminin isoforms promote attachment of hepatocytes via different integrins". *Exp. Cell Res.* 215, 33-39

42. Fässler, R., Pfaff, M., Murphy, J., Noegel, A. A., **Johansson, S.** Timpl, R., Albrecht, R. 1995. "Lack of β_1 integrin gene in embryonic stem cells affects morphology, adhesion and migration but not integration into the inner cell mass". *J. Cell Biol.* 128, 979-988

43. Prasthofer, T., Ek, B., Ekman, P., Hook, M., **Johansson, S.** 1995. "Protein kinase C phosphorylats two of the four known syndecan cytoplasmic domains *in vitro*". *Biochem. and Mol. Biol. Int.* 36, 793-802.

44. Holmvall. K., Camper, L., **Johansson, S.**, Kimura, J. H., Lundgren-Åkerlund, E. 1995. "Chondrocyte and chondrosarcoma cell integrins with affinity for collagen type II and their response to mechanical stress". *Exp. Cell Res.* 221, 496-503

45. Hauzenberger, D., Martin, N., **Johansson, S.**, Sundqvist, K.-G. 1996. "Characterization of lymphocyte fibronectin". *Exp. Cell Res.* 222, 312-318

46. Wennerberg, K., Lohikangas, L., Gullberg, D., Pfaff, M., **Johansson, S.**, Fässler, R. 1996. "Integrin beta1-dependent and -independent polymerization of fibronectin". *J. Cell Biol.* 132, 227-238

47. Frieser, M., Hallmann, R., **Johansson, S.**, Vestweber, D., Goodman, S. L., Sorokin, L. 1996. "Mouse polymorphonuclear granulocyte binding to extracellular matrix molecules involves β1 integrins". *Eur. J. Immunol.* 26, 3127-3136

48. Yu, J-L., **Johansson, S.**, Ljung, Å. 1997. "Fibronectin exposes different domaines after adsorption to a heparinized and an unheparinized polyvinyl chloride surface". *Biomaterials*, 18, 421-427.

49. Svineng, G., Fässler, R., **Johansson, S.** 1998. "Identification of β1C-2, a novel splice variant of the integrin subunit β1". *Biochem. J.* 330, 1255-1263.

50. Wennerberg, K., Fässler, R., Wärmegård, B., **Johansson, S.** 1998. "Mutational analysis of the potential phosphorylation sites in the cytoplasmic domain of integrin β1A. Requirement for threonines 788-789 in receptor activation". *J. Cell Sci.* 111, 1117-1126

51. Hirsch, E., Lohikangas, L., Gullberg, D., **Johansson, S.**, Fässler, R. 1998. "β1 is not essential for skeletal myogenesis in vivo but plays a role in vitro". *J. Cell Sci.*111, 2397-2409.

52. Brakebusch, C., Wennerberg, K., Krell, H. W., Weidle, U. H., Sallmyr, A., **Johansson, S.**, Fässler, R. 1999. " β1 integrin promotes but is not essential for metastsis of ras-myc transformed fibroblasts". *Oncogene*, 18, 3852-3861.

53. Su, B., **Johansson, S.**, Fällman, M., Patarroyo, M., Granstrom, M., Normark, S. 1999. "Signal transduction-mediated adherence and entry of Helicobacter pylori into cultured cells". *Gastroenterology*, 117, 595-604.

54. McCourt, P. A. G., Smedsrød, B., Melkko, J., **Johansson, S.** 1999. "Characterization of the hyaluronan receptor on liver endothelial cells and its functional relationship to the scavanger receptor." *Hepatology*, 30, 1276-1286.

55. Armulik, A., Nilsson, I., von Heijne, G., **Johansson, S.** 1999. "Delineation of the border between the transmembrane and cytoplasmic domains of human integrin subunits". *J. Biol. Chem.*, 274, 37030-37034.

56. Svineng, G., **Johansson, S.** 1999. "Integrin subunits β1C-1 and β1C-2 expressed in GD25 cells are retained and degraded intracellularly rather than localized to the cell surface". *J. Cell Sci.*, 112, 4751-4761.

57. Armulik, A., Svineng, G., Wennerberg, K., Fässler, R., **Johansson, S.** 2000. "Expression of integrin subunit β1B in β1-deficient GD25 cells does not interfere with αvβ1 functions". *Exp. Cell Res.*, 254, 55-63.

58. Wennerberg, K., Armulik, A., Sakai, T., Karlsson, M., Fässler, R., Schaefer, E. M., Mosher, D., **Johansson, S.** 2000. "The cytoplasmic tyrosines of integrin subunit β1 are involved in FAK activation". *Mol. Cell Biol.* 20, 5758-5765.

59. Fowler, T, Wann, E.R., Joh, D, **Johansson, S.**, Foster, T.J., Höök, M. 2000. "Staphylococcus aureus invasion of mammalian cells involves a fibronectin bridge between the fibronectin-binding MSCRAMMs and host β1 integrins". *Eur. J. Cell Biol.*79, 672-679.

60. Lohikangas, L., Gullberg, D., **Johansson, S.** 2001. "Assembly of laminin polymers is dependent on β1 integrins". *Exp. Cell Res.* 265, 135-144.

61. Politz, O., Grachev, A., McCourt, P., Schledzewski, K., Guillot, P., Longati, P., Johansson, S., Birk, P., Hakiy, N., Franke, P., Kodelja, V., Kannicht, C.,, Orfanos, C., **Johansson, S.**, Goerdt, S. 2002. "Stabilin-1 and stabiline-2 constitute a novel family of fasciclin-like hyaluronan receptor homologues". *Biochem. J.* 362,155-164.

62. Gustavsson, A., Armulik. A., Brakebusch, C., **Johansson, S.**, Fällman, M. 2002. "Role of the β1 integrin cytoplasmic tail in mediating invasin-promoted internalization of bacteria". *J. Cell Sci.* 115, 2669-2678.

63. Velling, T., Risteli, J., Wennerberg, K., Mosher, D.F., **Johansson S.** 2002. "Polymerization of Type I and III Collagens Is Dependent On Fibronectin and Enhanced By Integrins α11β1 and α2β1." *J. Biol. Chem.*, 277, 37377-37381.

64. Fowler, T., **Johansson, S.,** Wary, K. K., Hook M. 2003. "Src kinase inhibitors control in vitro cellular internalization of Staphylococcus aureus". *Cell. Microbiol.* 5, 417-426.

65. Armulik, A., Velling, T., **Johansson, S.** 2004. " The integrin □1 subunit transmembrane domain regulates PI3K-dependent tyrosine phosphorylation of CAS". *Mol. Biol. Cell*, 15, 2558-2567.

66. Stefansson, A,. Armulik, A., Nilsson, I-M. von Heijne, G., **Johansson**, S. 2004. "Determination of N- and C-terminal borders of the transmembrane domain of integrin subunits". *J. Biol. Chem.* 279, 21200-21205.
67, Velling, T., Nilsson, S., Stefansson, A., **Johansson S.** 2004. "β1-Integrins induce phosphorylation of PKB/Akt on serine 473 independently of focal adhesion kinase and Src family kinases". *EMBO Reports,* 5, 901-905.

68. Hansen B, Longati P, Elvevold K, Nedredal GI, Schledzewski K, Olsen R, Falkowski M, Kzhyshkowska J, Carlsson F, **Johansson S**, Smedsrød B, Goerdt S, McCourt P, Johansson S. (2005) "Stabilin-1 and stabilin-2 are both directed into the early endocytic pathway in hepatic sinusoidal endothelium via interactions with clahtrin/AP-2 independent of ligand binding". *Exp Cell Res.*, 303, 160-73.

69. Lannergård J, Flock M, **Johansson S**, Flock J-I , Guss B. 2005 " Studies of fibronectin-binding proteins of *Streptococcus equ*" *Infect. Immun.* 75, 7243-7251

70, Nilsson, S., Caniovska, D., Brakebush, C., Fässler, R., **Johansson S**. 2006. "Threonine-788 in integrin subunit β1 regulates integrin activation." *Exp Cell Res.*, 312, 844-853.

71. Velling, T, Stefansson, A.,. **Johansson S**. 2007. "EGFR and β1-Integrins utilise different signalling pathways to activate Akt" *Exp. Cell Res., in press.*

APPENDIX 2

OF
DECLARATION UNDER 37 C.F.R. § 1.132

DATED JANUARY 16, 2008

BY
STAFFAN JOHANSSON

Content:

Hynes, R.O., Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992 Apr 3;69(1):11-25. (15 pages)

Hynes, R.O., Cell adhesion: old and new questions. Trends Cell Biol. 1999 Dec;9(12):M33-7. (5 pages)

Loftus, J.C., et al., Integrin-mediated cell adhesion: the extracellular face. J Biol Chem. 1994 Oct 14;269(41):25235-8. (4 pages)

Gale, C., et al., Cloning and expression of a gene encoding an integrin-like protein in Candida albicans. Proc Natl Acad Sci USA. 1996 Jan 9;93(1):357-61. (5 pages)

Berg, R.W., et al., Cloning and characterization of a novel beta integrin-related cDNA coding for the protein TIED ("ten beta integrin EGF-like repeat domains") that maps to chromosome band 13q33: A divergent stand-alone integrin stalk structure. Genomics. 1999 Mar 1;56(2):169-78. (10 pages)

Laval, V., et al., A family of Arabidopsis plasma membrane receptors presenting animal beta-integrin domains. Biochim Biophys Acta. 1999 Nov 16;1435(1-2):61-70. (10 pages)

Rubin, G.M., et al., Comparative genomics of the eukaryotes. Science. 2000 Mar 24;287(5461):2204-15. (12 pages)

Forsberg, E., et al., Purification and characterization of integrin alpha 9 beta 1. Exp Cell Res. 1994 Jul;213(1):183-90. (8 pages)

Munger, J.S., et al., The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell. 1999 Feb 5;96(3):319-28. (10 pages)

Cell, Vol. 69, 11–25, April 3, 1992, Copyright © 1992 by Cell Press

Integrins: Versatility, Modulation, and Signaling in Cell Adhesion

Review

Richard O. Hynes
Howard Hughes Medical Institute and
Center for Cancer Research
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

The recognition of integrins as a widely expressed family of cell surface adhesion receptors is around five years old (Hynes, 1987). At that time, one could identify about ten distinct vertebrate integrins; there are now about twenty, and the number is still rising. Integrins appear to be the major receptors by which cells attach to extracellular matrices, and some integrins also mediate important cell–cell adhesion events. Through these functions they play important roles both in development and in adult organisms. Several human genetic diseases affecting integrins demonstrate their importance in various physiological and pathological processes, and the ability to interfere with integrin functions using antibodies or peptides offers many opportunities for therapeutic intervention in diseases as diverse as thrombosis, inflammation, and cancer. Because of these multifarious roles, integrins have been intensively studied by scientists in many different fields, and more than one integrin paper a day is now published. Despite this plethora of information, it is possible to discern common principles, and, in this brief review, I will attempt some generalizations and syntheses to make the field accessible to the nonspecialist. In particular, I will focus on recent evidence concerning regulation of integrin affinities and signaling events mediated by integrins.

Multiple Integrins and Multiple Ligands

All integrins are αβ heterodimers. The α subunits vary in size between 120 and 180 kd and are each noncovalently associated with a β subunit (90–110 kd). Most integrins are expressed on a wide variety of cells, and most cells express several integrins. The table summarizes the diversity of vertebrate integrins as currently understood. There are 8 known β subunits and 14 known α subunits, all of which have been sequenced at the cDNA level except α_7 and α_{IEL}. References for most of the subunits are given in earlier reviews (Albelda and Buck, 1990; Arnaout, 1990; Hemler, 1990; Springer, 1990a, 1990b; Ruoslahti, 1991).

Although the nomenclature given in Table 1 is the most widely used, some earlier names still persist in the literature. The platelet-specific integrin, $\alpha_{IIb}\beta_3$, is often referred to as GPIIb–IIIa (Kieffer and Phillips, 1990; Phillips et al., 1991), and other integrins expressed on platelets are sometimes given names based on earlier platelet glycoprotein nomenclature (GPIaIIa = $\alpha_2\beta_1$, GPIcIIa = $\alpha_5\beta_1$ plus $\alpha_6\beta_1$). The leukocyte-specific β_2 integrins are still referred to by earlier names ($\alpha_L\beta_2$ = LFA-1; $\alpha_M\beta_2$ = Mac-1, Mo-1, or CR3; $\alpha_X\beta_2$ = p150,95); these are the only integrins for which CD nomenclature is frequently used (β_2 = CD18; associated α subunits = CD11a,b,c). Some other integrin subunits also have assigned CD numbers (Hemler, 1990; Springer, 1990a, 1990b), but these are rarely used. Finally, several β_1 integrins are sometimes referred to as VLA (very late after activation) antigens, a name that arose from the time at which $\alpha_1\beta_1$ and $\alpha_2\beta_1$ appear on lymphocytes (Hemler, 1990). However, since most cells in the body express one or more β_1 integrins constitutively, and since some of the α subunits occur in association with other β subunits, the αβ nomenclature is more widely applicable.

Although 8 β subunits and 14 α subunits could in theory associate to give more than 100 integrin heterodimers, the actual diversity appears to be much more restricted. Many α subunits can associate with only a single β subunit. For instance, white blood cells express both β_1 and β_2 integrins, but each α subunit associates with only one of the two β subunits. Thus, subfamilies with shared β subunits can be defined (see table). However, several α subunits (α_4, α_6, α_v, and perhaps others) can associate with more than one β subunit; α_v is particularly promiscuous in this respect. The discovery of four new β subunits (Ramaswamy and Hemler, 1990; Suzuki et al., 1990; McLean et al., 1990; Sheppard et al., 1990; Yuan et al., 1990, 1991; Erle et al., 1991; Moyle et al., 1991) and two new α subunits (Kramer et al., 1991; von der Mark et al., 1991; Bossy et al., 1991) in the past two years suggests that others may be found as new techniques (e.g., polymerase chain reaction) are increasingly applied, and newly discovered subunits may turn out to have additional associations not yet recognized.

A further level of complexity is introduced by the existence of alternative splicing. In mammals, several subunits have alternatively spliced cytoplasmic domains. These include β_1 (Altruda et al., 1990), β_3 (van Kuppevelt et al., 1989), β_4 (Suzuki and Naitoh, 1990; Hogervorst et al., 1990; Tamura et al., 1990), α_3 (Tamura et al., 1991; C. M. DiPersio and R. Hynes, unpublished data), and α_6 (Hogervorst et al., 1991; Cooper et al., 1991). Human α_{IIb} can be alternatively spliced in the extracellular domain close to the membrane (Bray et al., 1990), the Drosophila PS2 α subunit can be alternatively spliced in a region adjacent to the metal-binding sites (Brown et al., 1989), and the PS3 β subunit can be alternatively spliced in the ligand-binding domain (G. Yee, R. Patel-King, F. Chen, and R. Hynes, unpublished data). Possible implications of the alternative splicing will be discussed later.

The functions (ligand and adhesive specificity) of individual integrins have been elucidated using cell adhesion assays, monoclonal antibodies, and affinity chromatography. Examination of Table 1 quickly reveals that individual integrins can often bind to more than one ligand. Equally, individual ligands are, more often than not, recognized by more than one integrin. Thus, earlier designations of integrins as fibronectin or vitronectin receptors have proven too restrictive—there are several receptors for each of these proteins, and many of them are not highly specific for individual adhesive ligands. The majority of the

Table 1. The Integrin Receptor Family

Subunits		Ligands and Counterreceptors	Binding Site
β_1[a]	α_1	Collagens, laminin	
	α_2	Collagens, laminin	DGEA[h]
	α_3[a]	Fibronectin, laminin, collagens[c]	RGD ± ?[i]
	α_4	Fibronectin (V25), VCAM-1	EILDV[k]
	α_5	Fibronectin (RGD)	RGD
	α_6[a]	Laminin	
	α_7	Laminin	
	α_8	?	
	α_v	Vitronectin, fibronectin (?)[d]	RGD
β_2	α_L	ICAM-1, ICAM-2	
	α_M	C3b component of complement (inactivated), fibrinogen, factor X, ICAM-1	
	α_X	Fibrinogen, C3b component of complement (inactivated)?	GPRP
β_3[a]	α_{IIb}	Fibrinogen, fibronectin, von Willebrand factor, vitronectin, thrombospondin	RGD, KQAGDV[l]
	α_v	Vitronectin, fibrinogen, von Willebrand factor, thrombospondin, fibronectin, osteopontin, collagen	RGD
β_4[a]	α_6[a]	Laminin ??[e]	
β_5	α_v	Vitronectin	RGD
β_6	α_v	Fibronectin[f]	RGD
β_7(= β_P?)	α_4	Fibronectin (V25), VCAM-1[g]	EILDV[k]
	α_{IEL}[b]	?	
β_8	α_v	?	

The current spectrum and interactions of vertebrate integrins are listed. cDNA sequences for all subunits except α_7 and α_{IEL} have been reported (see data banks or earlier reviews for references). Several subfamilies exist, each with 2–9 α subunits and a common shared β subunit (β_1, β_2, β_3, or β_7). In addition, several of the α subunits can interact with other β subunits (β_4–β_8). Each αβ receptor recognizes one or more extracellular ligands or counterreceptors on other cells. It should be noted that the ligand specificity of a given receptor can be markedly affected by its environment or state of activation (see text). The peptide recognition sequences in the ligands are given where known; where none is given, the evidence indicates that the recognition sequence is *not* RGD.

[a] These subunits can have alternatively spliced cytoplasmic domains.

[b] The subunit designated α_{IEL} is expressed specifically on intraepithelial lymphocytes (IEL) in association with β_7 (Yuan et al., 1991; Parker et al., 1992). However, α_E has previously been used for α_6 in association with β_4 (Kajiji et al., 1989), so this term has not been used here to avoid confusion.

[c] A recent paper (Carter et al., 1991) describes a ligand designated epiligrin that is not fully characterized.

[d] Publications differ as to the specificity of this receptor (Bodary and McClean, 1990; Vogel et al., 1990).

[e] The specificity of this receptor is controversial (Lotz et al., 1990; Sonnenberg et al., 1990).

[f] Busk et al. (1992).

[g] The specificity of this receptor is not yet clear. $\alpha_4\beta_7$ on some cells can apparently bind the same ligands as $\alpha_4\beta_1$ (Rüegg et al., 1992), whereas on others it requires activation and is much less effective than $\alpha_4\beta_1$ (Chan et al., 1992a).

[h] Defined in type I collagen only (Staatz et al., 1991).

[i] Publications differ as to whether or not this receptor recognizes RGD (Wayner et al., 1988; Hynes et al., 1989; Elices et al., 1991).

[k] Defined only in the alternatively spliced V segment of fibronectin.

[l] RGD is recognized in all ligands; KQAGDV is recognized only in fibrinogen γ chain.

ligands listed in the table are extracellular matrix proteins involved in cell–substratum adhesion. However, some of these, such as fibrinogen, can also mediate cell–cell aggregation, and some integrins recognize integral membrane proteins of the immunoglobulin superfamily (ICAM-1, ICAM-2, VCAM-1) and mediate direct cell–cell adhesion.

Considerable progress has been made in defining the integrin recognition sites in the ligands and counterreceptors listed in the table. The first binding site to be defined was the Arg-Gly-Asp (RGD) sequence present in fibronectin, vitronectin, and a variety of other adhesive proteins. This tripeptide sequence is recognized by several integrins ($\alpha_5\beta_1$, $\alpha_{IIb}\beta_3$, and all or most $\alpha_v\beta$ integrins) but not by most others. $\alpha_{IIb}\beta_3$ recognizes, in addition, the sequence Lys-Gln-Ala-Gly-Asp-Val (KQAGDV) in fibrinogen. Other integrins recognize different sequences: $\alpha_2\beta_1$ binds Asp-Gly-Glu-Ala (DGEA) in type I collagen, $\alpha_4\beta_1$ binds Glu-Ile-Leu-Asp-Val (EILDV) in an alternatively spliced segment of fibronectin, and it has recently been reported that $\alpha_X\beta_2$ binds Gly-Pro-Arg-Pro (GPRP) in fibrinogen (Loike et al., 1991). Other binding sites have not yet been defined as precisely, although the various laminin receptors recognize specific parts of the laminin molecule (e.g., Hall et al., 1990), and those integrins binding immunoglobulin superfamily counterreceptors recognize specific immunoglobulin-like domains (Staunton et al., 1990; Diamond et al., 1991).

Figure 1 summarizes some of the relationships and interactions of individual integrin subunits. It is based on a dendrogram relating the sequences of various α subunits. The sequence homologies allow one to cluster the α subunits in several subgroups, and Figure 1 also depicts other relationships that reinforce this conclusion. Thus, all the RGD-reactive integrins are relatively closely related, and all contain cleaved α subunits without I domains (see be-

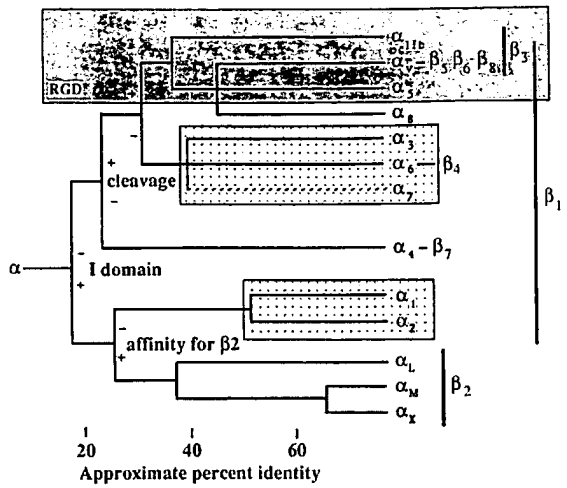


Figure 1. Structural, Functional, and Evolutionary Relationships among Integrins

The figure depicts schematically the sequence similarities among different human α subunits. Also indicated are various structural and functional features, such as whether or not α subunits are posttranslationally cleaved, the presence or absence of I domains (homologous segments of 180 amino acids inserted in the extracellular domains of some α subunits), and interactions of individual α subunits with various β subunits. All α subunits except α_{IIb} and the three β_3-associated subunits can bind to β_1; only α_v can bind to more than two β subunits. Shaded box indicates the subset of related integrins that recognize RGD sequences, and the stippled boxes indicate two distantly related groups of laminin-binding integrins. (α_7 is not yet fully sequenced, but its N-terminal sequence is most closely related to that of α_6.) The two sets of laminin-binding integrins appear to recognize different parts of the laminin molecule, suggesting that laminin recognition may have evolved twice. The gene structures of α_{IIb} (Heidenreich et al., 1990) and α_v (Corbi et al., 1990) are related to each other and to that of Drosophila PS2 α (Brown et al., 1989), confirming the ancient evolutionary origin of α subunits.

low). There are two separate clusters of laminin receptors. The members of the first group ($\alpha_3\beta_1$, $\alpha_6\beta_1$, and $\alpha_7\beta_1$) all recognize the long arm of laminin, and all contain cleaved α subunits. The second group ($\alpha_1\beta_1$ and $\alpha_2\beta_1$) is distantly related to the first group, although α_1 and α_2 are very similar to each other (50% identity, and both are uncleaved α subunits with I domains). $\alpha_1\beta_1$ recognizes the cross region of laminin; it is a reasonable prediction that $\alpha_2\beta_1$ binds to the same region. It appears that laminin recognition has evolved twice in different integrin subsets. Similarly, recognition of two different sequences in fibronectin apparently has evolved twice ($\alpha_5\beta_1$ recognizes RGD, and $\alpha_4\beta_1$ EILDV), as has recognition of different parts of fibrinogen by β_2 and β_3 integrins.

Since the sequences and genes of Drosophila integrins are about as closely related as those of the most divergent vertebrate subunits, it is clear that integrins arose at a very early point in evolution, before divergence of the protostome and deuterostome lineages. Indeed, one can argue that metazoans would require integrins or something analogous to maintain multicellularity. It is clear that subsequent divergence of the integrin subunits has allowed development of great versatility in the cell adhesion mediated by these receptors.

Transmembrane Topography and Interactions

Both subunits of integrins are transmembrane glycoproteins, each with a single hydrophobic transmembrane segment (Figure 2). In most integrins, the cytoplasmic domains are short (50 amino acids or less). β_4 is a notable exception: its cytoplasmic domain comprises over 1000 amino acids. The extracellular domains (>75 kd for β subunits, and >100 kd for α subunits) associate to form the αβ heterodimers. Truncated forms lacking transmembrane and cytoplasmic domains can be expressed and do form functional αβ dimers (Dana et al., 1991; Bodary et al., 1991), indicating that αβ interactions do not rely on the transmembrane or cytoplasmic domains; this conclusion is also supported by the selective α associations of chimeric β subunits (Solowska et al., 1991). Electron microscopic images of several integrins show a globular head apparently comprising parts of both subunits and two stalks extending to the lipid bilayer (Carrell et al., 1985; Kelly et al., 1987; Nermut et al., 1988). Both subunits contain extensive disulfide bonding, the patterns of which have been partially elucidated (Calvete et al., 1989, 1991). Consistent with a model of compact folded domains, integrins are fairly resistant to proteolysis of intact cells.

These and other observations give rise to the models depicted in Figure 2. Characteristic of all β subunits is a four-fold repeat of a cystine-rich segment believed to be internally disulfide bonded (Calvete et al., 1991). The N-terminal 40–50 kd is tightly folded with internal disulfide loops and contributes to the ligand-binding domain (see below). The α subunits all contain a seven-fold repeat of a homologous segment; the last three or four of these repeats contain sequences (Asp-x-Asp-x-Asp-Gly-x-x-Asp or related sequences) that likely contribute the divalent cation–binding properties of these subunits. Divalent cations are essential for receptor function. The nature of the cations can affect both affinity and specificity for ligands and divalent cations are necessary for αβ subunit associations of some integrins (Gailit and Ruoslahti, 1988; Kirchhofer et al., 1990a, 1991). This part of the α subunit also contributes to the ligand-binding domain (see below).

Some α subunits are posttranslationally cleaved to give a 25–30 kd transmembrane chain disulfide-bonded to a larger, wholly extracellular chain (Figure 2). Other α subunits contain an extra segment of around 180 amino acids, known as an I domain, which is inserted before the last five homologous repeats containing the cation-binding domains. The functions of the I domains are unknown, but they are homologous to collagen-binding domains of von Willebrand factor and to cartilage matrix protein and complement proteins. I domains are characteristic of β_2–associated α subunits and of the α_1 and α_2 subunits, which contribute to the collagen–laminin receptors $\alpha_1\beta_1$ and $\alpha_2\beta_2$ (Figure 1). The best current guess is that the I domains contribute ligand-binding functions to these integrins, but that remains to be proven. Interestingly, one of the Drosophila integrin α subunits contains an alternatively spliced segment that can be inserted just N-terminal to the cation-binding domains (Brown et al., 1989). This is close to the position of the I domains and may also affect

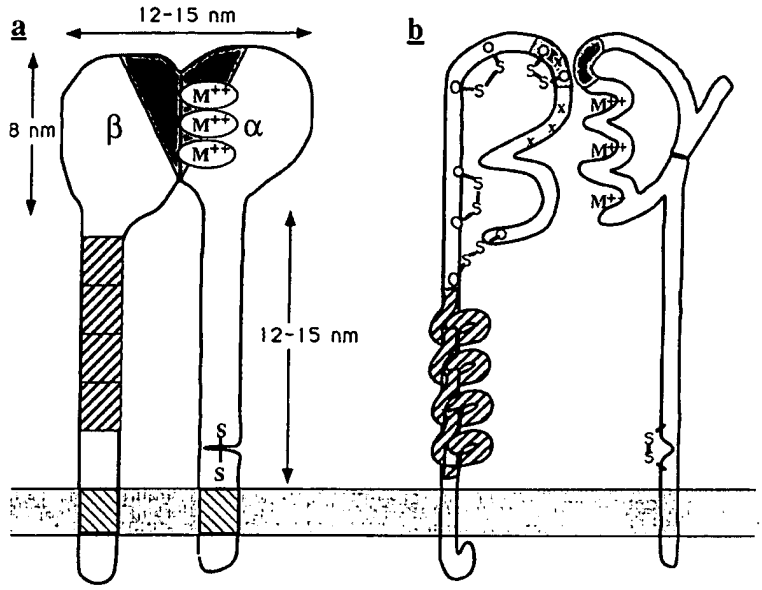


Figure 2. Structural Features of Integrin Receptors

(a) shows the overall shape, as deduced from electron microscopy, as well as the putative locations of the cystine-rich repeats of the β subunit (crosshatched) and metal-binding sites in the α subunit (M^{++}). The shaded area represents the ligand-binding region that is known to be made up from both subunits based on cross-linking and binding data.

(b) schematizes the arrangement of the polypeptide chains with the cystine repeats internally folded and the head region of the β subunit containing internal disulfide loops, some but not all of which are shown. A disulfide bond from the middle of the β subunit to a point close to the membrane has been proposed (Calvete et al., 1991) but is omitted here for clarity. Xs indicate positions of mutations (of human β_2 or β_3 subunits) known to affect ligand binding or αβ dimerization. The positions of alternatively spliced segments in Drosophila subunits are shaded.

ligand-binding specificity or β subunit associations (see also Figure 2 and below).There is also alternative splicing in the ligand-binding domain of a Drosophila β subunit (G. Yee, R. S. Patel-King, F. Chen, and R. O. Hynes, unpublished data). So far, no such splicing of ligand-binding domains has been detected in vertebrates.

The position of the ligand-binding site(s) within the integrin subunits can be deduced from several lines of evidence. Chemical cross-linking data on the two β_3 integrins, using peptides bound by these integrins, places the ligand in proximity both to the divalent cation–binding domains of the α subunits and to the segment from approximately residues 100–200 of the β_3 subunit (D'Souza et al., 1988, 1990; Smith and Cheresh, 1988, 1990). A point mutation at position 119 of the β_3 subunit ablates ligand binding (Loftus et al., 1990), and several mutations in the corresponding region of the β_2 subunit affect $\alpha\beta_2$ associations (Kishimoto et al., 1989; Wardlaw et al., 1990; Arnaout et al., 1990). These and other data are consistent with the models shown in Figure 2, in which both α and β subunits contribute to the ligand-binding site that lies at or near the interface between the two subunits. This is consistent with the observation (see table) that switching either the α or the β subunits in integrins can lead to changes in ligand specificity. As mentioned earlier, it is of interest that the I domains, when present, and the alternative splicing events detected in Drosophila integrin subunits all fall in the same regions of the receptors.

Therefore, the current picture is that the N-terminal domains of α and β subunits combine to form a ligand-binding head on each integrin. This head is connected by two stalks, each made up of one of the two subunits, to the membrane-spanning segments and thus to the two cytoplasmic domains. These cytoplasmic domains are believed to interact with cytoskeletal proteins and perhaps with other cytoplasmic components. The evidence for cytoskeletal connections comes from a variety of sources and includes light and electron microscopic evidence for colocalization of integrins and cytoskeletal structures (see Burridge et al., 1988, for review), fluorescence photobleaching evidence for restricted mobility of integrins in focal contacts, which are points of cell–substratum and cytoskeleton–membrane contact (Duband et al., 1986), and biochemical evidence for interactions of integrins or cytoplasmic domain peptides with the cytoskeletal proteins, talin (Horwitz et al., 1986; Tapley et al., 1989) and α-actinin (Otey et al., 1990). Deletion of all or part of the β_1 cytoplasmic domain interferes with associations with focal contacts (Solowska et al., 1989; Hayashi et al., 1990; Marcantonio et al., 1990).

There is less direct evidence for interactions of α subunit cytoplasmic domains with the cytoskeleton, but it is of interest that different α subunits have very different cytoplasmic sequences, and that different receptors for a given ligand can differ in their apparent associations with the cytoskeleton. For example, $\alpha_5\beta_1$ is in focal contacts, whereas $\alpha_3\beta_1$ is not, even though both interact with fibronectin (Elices et al., 1991). Similarly, $\alpha_v\beta_3$ is in focal contacts, whereas $\alpha_v\beta_5$ is not, even when both interact with the RGD site in vitronectin (Wayner et al., 1991). Such results indicate the importance of different integrin subunits in mediating differing cellular responses to common extracellular ligands. A recent paper analyzing chimeric integrin α subunits shows that different cytoplasmic domains trigger different functions (collagen gel contraction or migration) when transfected into cells (Chan et al., 1992b). The existence of alternative cytoplasmic domains on several integrin subunits presumably contributes further versatility although there is as yet no evidence concerning their functions.

Although most integrins are thought to interact somehow with the actin-based cytoskeleton, $\alpha_6\beta_4$ clearly plays

a different role. This integrin with its large β_4 cytoplasmic domain is specifically concentrated at hemidesmosomes in epithelial cells (Stepp et al., 1990; Sonnenberg et al., 1991; Kurpakus et al., 1991), where it most likely interacts somehow with intermediate filaments, which are characteristically associated with hemidesmosomes.

The details and subtleties of integrin–cytoskeleton associations need much further study, but it seems clear that one major function of integrins is to mediate cytoskeletal interactions at the inner face of the membrane at sites of cell–substratum or cell–cell adhesion. I will return later to the possibility that integrins may also transmit other transmembrane signals and to the general question of how extracellular ligand occupancy is coupled to intracellular events.

Modulation of Integrin Affinities and Specificities

Given the wide variety of integrins, individual cells can and do vary their adhesive properties by selective expression of integrins. Further versatility is introduced by the ability of cells to modulate the binding properties of integrins. The ligand specificities of different integrins shown in the table are those that can be demonstrated for each receptor. However, in a given cell, a particular integrin may not exhibit all the specificities listed. For instance, $\alpha_2\beta_1$ on platelets is specific for collagen and not laminin (Staatz et al., 1989), whereas on other cells it can recognize both ligands (Elices and Hemler, 1989; Kirchhofer et al., 1990b). While the possibility is not completely ruled out that this particular difference could reflect variant forms of $\alpha_2\beta_1$ (e.g., alternatively spliced forms), there are other instances in which a purified integrin displays different ligand specificities depending on context (lipids [Conforti et al., 1990] or divalent cations [Kirchhofer et al., 1990a, 1991]).

Perhaps more significantly, the specificity and affinity of a given integrin receptor on a given cell are not always constant. There are numerous examples of modulation of integrin function. Both activation and deactivation of integrin functions have been reported. The best understood examples are $\alpha_{IIb}\beta_3$ on platelets (Kieffer and Phillips, 1990; Phillips et al., 1991) and β_2 integrins on neutrophils, monocytes, and lymphocytes. These will serve as a basis to discuss the general principles that very likely apply to many other instances of integrin modulation.

Integrin $\alpha_{IIb}\beta_3$ on resting circulating platelets does not bind any of its soluble ligands, which is a good thing, since such binding would lead to thrombosis. Unactivated platelets bind to surface-bound fibrinogen via $\alpha_{IIb}\beta_3$ and can thus join hemostatic events already underway. However, only after platelet activation by thrombin, collagen, or other platelet agonists does $\alpha_{IIb}\beta_3$ become an effective receptor for soluble fibrinogen or the other ligands listed in the table (Kieffer and Phillips, 1990; Phillips et al., 1991). This activation event has been known for a long time and is still not fully understood. Activation is accompanied by a conformational change in the $\alpha_{IIb}\beta_3$ receptor that can be detected immunologically (Shattil et al., 1985; Gulino et al., 1990; Kouns et al., 1990; O'Toole et al., 1991a; Andrieux et al., 1991) or biophysically (Parise et al., 1987; Sims et al., 1991). A further conformational change occurs on ligand binding (Frelinger et al., 1988; 1990; 1991).

Activation of $\alpha_{IIb}\beta_3$ can be accomplished by activation of the platelets, an event that involves activation of several G proteins, increases in intracellular pH and Ca^{++}, phosphatidyl inositol turnover, and activation of protein kinases (Manning and Brass, 1991; Shattil and Brugge, 1991). Receptor activation can also be produced by certain antibodies against the receptor, either in intact cells or with solubilized receptor (O'Toole et al., 1991a; Gulino et al., 1990; Kouns et al., 1990; Frelinger et al., 1991). Interestingly, activation can also be accomplished by the ligands themselves (Du et al., 1991).

All these results are consistent with a conformational switch (or switches) between states of the extracellular domain of $\alpha_{IIb}\beta_3$, normally driven from within the cell after its activation. Such a model would also predict that the activated state would be favored by interactions with molecules that bind to the activated state of the extracellular domain (e.g., antibodies, ligands). How could such a conformational change be driven from within the cell? Some clues come from recent results using recombinant DNA expression methods. $\alpha_{IIb}\beta_3$ expressed in heterologous cells (O'Toole et al., 1991a; Kieffer et al., 1991) is in the resting or unactivated state. The activated state can be induced by monoclonal antibodies; even solubilized receptor can be activated by these monoclonal antibodies. Thus, the two states are intrinsic to the receptor itself.

O'Toole et al. (1991b) have shown that deletion of the cytoplasmic domain of α_{IIb} leads to a receptor that is constitutively active. Interestingly, substitution of the α_5 cytoplasmic domain did not restore the unactivated state of the receptor. Thus, the cytoplasmic domain of α_{IIb} in some way controls the binding affinity of the extracellular domain of $\alpha_{IIb}\beta_3$, maintaining the unactivated state. In activated platelets, this control is lifted. How this is achieved remains unclear, but the various second messenger pathways triggered in activated platelets suggest phosphorylation (of the integrin or of associated proteins) or lipid mediators as candidates. While phosphorylation of $\alpha_{IIb}\beta_3$ has been reported, it is of low stoichiometry and uncertain significance (Hillery et al., 1991; Shattil and Brugge, 1991).

The β_2 integrins (reviewed in Arnaout, 1990; Larson and Springer, 1990) expressed on leukocytes exhibit activation phenomena that are strikingly similar to those shown by $\alpha_{IIb}\beta_3$. Activation of leukocytes is required for expression of the various ligand-binding activities of the β_2 integrins (see table). Activation is accompanied by a conformational change(s) in the β_2 integrins that can be detected by specific monoclonal antibodies, including one, 7E3, originally isolated by its ability to recognize activated $\alpha_{IIb}\beta_3$ (Altieri and Edgington, 1988; Coller, 1985). Furthermore, at least one of these monoclonal antibodies (NKI-L16) "activates" the functions of $\alpha_L\beta_2$ (Keizer et al., 1988; van Kooyk et al., 1991). This monoclonal appears to be recognizing an activated state of the receptor and stabilizing it. Just as discussed earlier for $\alpha_{IIb}\beta_3$, all the data are consistent with switches between unactivated and activated conformations of the β_2 receptors.

The effective activation stimuli for β_2 integrins vary de-

pending on cell type. The earliest observations were on monocytes, in which $\alpha_M\beta_2$ (complement receptor CR3) can be activated by phorbol esters (Wright and Silverstein, 1982) or by adherence to fibronectin (Wright et al., 1983, 1984). These results presaged many more recent papers demonstrating two key features of integrin function: regulation from within the cell (inside-out signaling) and modulation of cellular behavior by extracellular matrix (outside-in signaling). Greatest progress in understanding the β_2 integrins has been made in neutrophils and lymphocytes, and I will concentrate on recent results on these two systems, which shed some light on the activation phenomena.

Neutrophils and monocytes need to attach to endothelial layers in order to leave the bloodstream at sites of inflammation. These extravasation events involve several different adhesive proteins (reviewed in Carlos and Harlan, 1990; Osborn, 1990). Central among these are the β_2 integrins—in the genetic disease leukocyte adhesion deficiency (LAD), β_2 integrins are absent, and leukocytes cannot adhere stably to the endothelium. In the normal process of arrest and extravasation, the first event is that the leukocytes roll along the vessel wall. This is mediated by adhesion receptors of the selectin family (Bevilacqua et al., 1991; Lasky and Rosen, 1992; Smith et al., 1991; von Andrian et al., 1991; Lawrence and Springer, 1991).

Selectin-mediated rolling is necessary but not sufficient for β_2 integrin–mediated adhesion. The latter requires activation of the integrins, which on circulating leukocytes are in their inactive state. Activation can be accomplished by phorbol esters or, more physiologically, by various inflammatory mediators (e.g., tumor necrosis factor, C5a, platelet activating factor, or fMet-Leu-Phe). Hermanowski-Vosatka et al. (1992) have recently described a lipid mediator purified from neutrophils stimulated by such agonists, which activates $\alpha_M\beta_2$ or $\alpha_L\beta_2$ either in cells or as purified receptors. Exactly how this lipid modulates integrin affinity is unclear, but its effects are reminiscent of earlier results showing that the specificity of purified $\alpha_v\beta_3$ for different ligands is modulated by the lipid composition of the liposomes used for the assays (Conforti et al., 1990).

Thus, the current view of leukocyte adhesion to endothelium at sites of inflammation invokes a multistep process involving initial unstable adhesion mediated by selectins (whose expression on the endothelium is induced by inflammation), followed by activation of leukocyte β_2 integrins by inflammatory mediators (produced by the endothelium or underlying inflamed tissue), and finally, strong adhesion of the β_2 integrins to counterreceptors (also induced on the endothelial cells). Thus, there is an adhesive cascade leading to function of the β_2 integrins only in the appropriate places: the integrins provide the necessary strong adhesion, but the specificity comes from the involvement of multiple receptors and, crucially, from the activation steps (Butcher, 1991; Hynes and Lander, 1992).

An analogous situation exists for T lymphocytes. Specificity of their adhesion to antigen-presenting cells comes from the T cell receptor, which recognizes antigenic peptides bound to major histocompatibility molecules. However, adhesion also relies on $\alpha_L\beta_2$ integrin (lymphocyte function antigen-1, or LFA-1) and can be blocked by antibodies to this integrin. $\alpha_L\beta_2$ binds to ICAM-1 on the target cells, but this is not an antigen-specific interaction. It turns out that cross-linking of either the CD3 component of the T cell receptor or of the costimulating receptor, CD2, activates $\alpha_L\beta_2$ on the T cells (Dustin and Springer, 1989; van Kooyk et al., 1989). Activation via CD3 is transient, allowing both adhesion and deadhesion. Thus, strong, antigen-specific adhesion is again mediated by an adhesion cascade: weak but specific adhesion via the T cell receptor–CD3 complex triggers activation of $\alpha_L\beta_2$, leading to strong adhesion. As in leukocytes, the integrin provides the adhesive strength, but the activation steps provide the specificity.

How is this activation accomplished? It is known that CD2 and CD3 cross-linking can activate protein kinase C in lymphocytes, and phorbol esters readily activate $\alpha_L\beta_2$ in lymphocytes (Dustin and Springer, 1989; van Kooyk et al., 1989), just as in leukocytes. These results suggest the involvement of protein kinase C in the activation pathways in both cell types. The transience of the activation mediated by T cell receptor–CD3 suggests that both activation and deactivation mechanisms exist, and treatments that elevate cAMP abrogate the T cell receptor–CD3-mediated activation of $\alpha_L\beta_2$ (Dustin and Springer, 1989).

Recombinant DNA expression experiments are providing some information about possible sites of regulation, although as for $\alpha_{IIb}\beta_3$, it is not yet possible to discern the details. In contrast with $\alpha_{IIb}\beta_3$, $\alpha_L\beta_2$ expressed in heterologous cells is constitutively active (Larson et al., 1991). When transfected into LAD mutant lymphoblastoid lines, recombinant $\alpha_L\beta_2$ exhibits phorbol-induced activation for binding to ICAM-1 (Hibbs et al., 1991a). In partial conformity with results on $\alpha_{IIb}\beta_3$, this binding activity is modulated by cytoplasmic domain sequences; deletions of the β_2 cytoplasmic domain lead to expression of inactive receptor (Hibbs et al., 1991a). These inactive receptors can be partially activated by the activating monoclonal antibody NKI-L16 or by phorbol esters. Truncations of the α subunit have no effect. Therefore, as for $\alpha_{IIb}\beta_3$, the cytoplasmic domains of β_2 integrins are targets for regulatory events. These could include phosphorylation or binding of some cytoplasmic component (cytoskeleton or other). Phorbol esters induce phosphorylation on serine residue(s) of the β_2 cytoplasmic domain in monocytes and neutrophils (Chatila et al., 1989; Buyon et al., 1990; Valmu et al., 1991), but there is no direct evidence that this event affects function. In fact, mutagenesis of the β_2 cytoplasmic domains can dissociate phosphorylation from activation in lymphoid cells (Hibbs et al., 1991b).

In addition to these relatively well-studied examples of integrin modulation, there are many other cases in which intracellular events affect the affinity of various integrins. Most are listed on the left-hand side of Figure 3 and will be mentioned briefly here. β_1 integrins are widely expressed on lymphocytes and leukocytes (Hemler, 1990; Shimizu and Shaw, 1991). As a generalization, the levels of β_1 integrins increase after antigen stimulation. As in other cell types, the β_1 integrins mediate attachment of lymphocytes to extracellular matrix proteins and likely play a role in extravasation and migration of activated lympho-

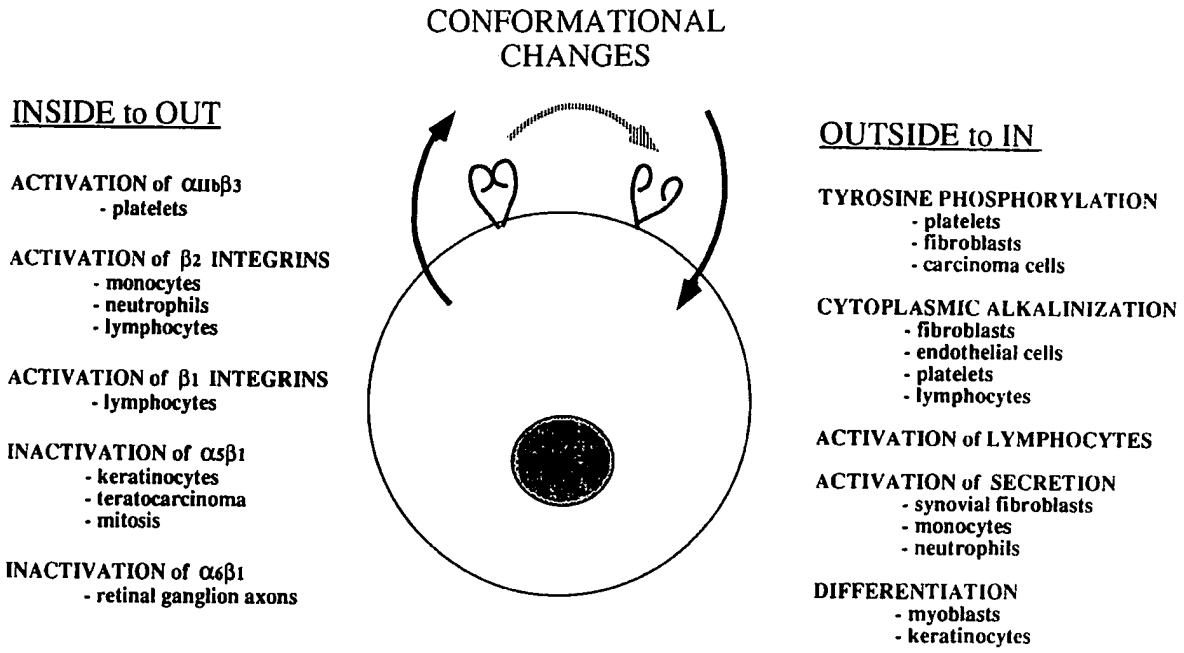


Figure 3. Signaling via Integrin Receptors

The receptors undergo conformational changes between at least two states: inactive (closed symbol) and active (open symbol). Only in the latter state do they bind most of their ligands. Signaling via integrins takes two forms: regulation of the affinity and conformation of the receptor from inside the cell (inside-out signaling), and triggering of intracellular events by ligand occupation of the receptors (outside-in signaling).

cytes in tissues during immune responses (e.g., Ferguson et al., 1991). $\alpha_4\beta_1$, which recognizes VCAM-1 on activated endothelial cells as well as an alternatively spliced segment of fibronectin, is involved in attachment of lymphocytes to endothelial layers under some conditions (B. R. Schwartz et al., 1990; Shimizu et al., 1991).

Activation of T cells by antigen or by phorbol esters leads to activation of $\alpha_2\beta_1$, $\alpha_4\beta_1$, $\alpha_5\beta_1$, and $\alpha_6\beta_1$ without changes in surface levels (Shimizu et al., 1990b; Chan et al., 1991; Wilkins et al., 1991). These changes in activity are over and above the changes in surface levels that occur on conversion from naive to memory cells and lead to increased adhesion of the activated T cells to collagen, fibronectin, and laminin. Therefore, it seems clear that β_1 integrins in lymphocytes undergo activation in a fashion similar to that described above for $\alpha_L\beta_2$.

In contrast with these activation events, there are several reported cases in which integrins lose activity during development but persist on the surface. This is so for $\alpha_5\beta_1$ in teratocarcinoma cells (Dahl and Grabel, 1989) and in keratinocytes (Adams and Watt, 1990) and for $\alpha_6\beta_1$ in retinal neurons (Neugebauer and Reichardt, 1991). Interestingly, loss of $\alpha_5\beta_1$ activity in teratocarcinoma cells correlates with loss of phosphoserine labeling of the integrin, although there is no evidence for causal linkage.

There are two reported instances in which phosphorylation of β_1 integrins is associated with apparent inactivation of the receptors. The first is during oncogenic transformation by Rous sarcoma virus; pp60^{src} phosphorylates a tyrosine residue in the β_1 subunit and this appears to reduce binding of β_1 integrins to both talin and fibronectin (Tapley et al., 1989). Phosphorylation of β_1 by pp60^{src} is absent in cells transformed by a virus mutant in pp60^{src} that fails to induce cell rounding and loss of fibronectin (Horvath et al., 1990), supporting the correlation of tyrosine phosphorylation of β_1 with inactivation. The second potential inactivating phosphorylation is of a serine two residues from the tyrosine phosphorylated by pp60^{src}. This serine becomes phosphorylated in mitotic cells and the $\alpha_5\beta_1$ from such cells no longer binds to fibronectin, consistent with the rounding and detachment of cells during mitosis (C. Grandori and R. Hynes, unpublished data).

Therefore, both activation and inactivation of various integrins can be mediated from within cells. The exact mechanisms are not fully elucidated, but presumptive evidence exists in several cases, implicating interactions with the cytoplasmic domains of integrin subunits. As mentioned earlier, the cytoplasmic domains of different integrins differ substantially in sequence and in several cases can be alternatively spliced. The potential modifications and interactions of these cytoplasmic domains with cytoskeletal and regulatory components should be a fruitful area for research in the next couple of years. So much for our current understanding of inside-out signaling via integrins. What about the possibility that integrin receptors can transmit signals into cells?

Do Integrins Transmit Signals into Cells?

There is increasing evidence that integrins do mediate information transfer into cells. The principal processes in which integrin-mediated signaling has been implicated are listed on the right-hand side of Figure 3.

In blood platelets, there is good evidence that tyrosine

phosphorylation events occurring upon platelet activation require ligand (i.e., fibrinogen) occupation of the major $\alpha_{IIb}\beta_3$ integrin for full expression (reviewed in Shattil and Brugge, 1991). Platelets can be activated by a variety of agonists, including thrombin, epinephrine, and ADP, all of which act via seven-transmembrane-helix receptors coupled to G proteins (Manning and Brass, 1991). Another platelet agonist, collagen, binds to an integrin receptor $\alpha_2\beta_1$ (Staatz et al., 1989). The second-messenger pathways activated by these and other agonists include several phospholipases, phosphatidyl turnover, elevation of cytoplasmic pH and Ca^{++}, and activation of protein kinases. Platelets contain several protein tyrosine kinases, and extensive tyrosine phosphorylation occurs on platelet activation (Shattil and Brugge, 1991, and references therein). Absence of $\alpha_{IIb}\beta_3$ or blockade of fibrinogen interaction with the activated receptor by antibodies or peptides interferes with the tyrosine phosphorylation events (Ferrell and Martin, 1989; Golden et al., 1990). Ligand binding of $\alpha_{IIb}\beta_3$ appears necessary but not sufficient for the phosphorylation response.

Several protein tyrosine kinases of the *src* family are associated with another adhesion receptor of platelets, GPIV or CD36, which is a receptor for thrombospondin and possibly collagen (Shattil and Brugge, 1991; Huang et al., 1991). During platelet aggregation, fibrinogen and thrombospondin interact, which could easily lead to coclustering of the two receptors in the plane of the membrane. The nature of the phosphorylated proteins and their role in subsequent activation events is unknown, but it seems reasonably clear that ligand binding of $\alpha_{IIb}\beta_3$ integrin, GPIV, and possibly $\alpha_2\beta_1$ can contribute to signal transduction events in platelets. One can perhaps view them as coreceptors, with the seven-transmembrane-helix receptors mediating the actions of soluble agonists of platelets.

Tyrosine phosphorylation events triggered via integrins have also been described in KB carcinoma cells and NIH 3T3 fibroblasts. Cross-linking of $\alpha_3\beta_1$ integrin in KB cells leads to transient tyrosine phosphorylation of protein(s) of 115–130 kd (Kornberg et al., 1991), and adhesion and spreading of NIH 3T3 cells on fibronectin or on anti-integrin antibodies leads to rapid tyrosine phosphorylation of a protein of similar size (Guan et al., 1991). Both are probably the same protein as one described by Kanner et al. (1990) as a tyrosine-phosphorylated protein in $pp60^{src}$-transformed cells. They may also be related to proteins of similar size that are phosphorylated after stimulation of cells with various soluble growth factors (Rees-Jones and Taylor, 1985; Sadoul et al., 1985; Pasquale et al., 1988). If, indeed, all these proteins are related, it would suggest convergence of the signaling pathways triggered by soluble growth factors, extracellular matrix adhesion receptors, and tyrosine kinase oncogene products. That, in turn, could offer potential explanations for the anchorage dependence of growth of normal cells and its loss in transformed cells (Guan et al., 1991).

The stimulation of tyrosine phosphorylation via integrins in platelets and adherent cells is of some interest, since tyrosine phosphate is found concentrated at cell–substratum and cell–cell contact points (Maher et al., 1985; Tsukita et al., 1991). Recognition of tyrosine-phosphate sites by proteins containing SH2 (*src* homology 2) domains (Koch et al., 1991) could contribute to the associations of structural and regulatory proteins into the submembranous cytoskeletal structures at such points of cell contact.

A second, well-established cytoplasmic event triggered via integrins is cytoplasmic alkalinization. Adhesion of fibroblasts, endothelial cells, and lymphocytes to fibronectin causes elevation of cytoplasmic pH, which correlates with the parallel stimulation of spreading and growth (Schwartz et al., 1989, 1991a; Ingber et al., 1990). Furthermore, by attaching the fibronectin or antibodies against α_5 or β_1 integrins to beads, it can be demonstrated that the effect on cytoplasmic pH is integrin mediated and can be uncoupled from effects on spreading or growth (Schwartz et al., 1991b). Since constitutively elevated cytoplasmic pH correlates with anchorage independence of growth in transformed cells (M. A. Schwartz et al., 1990), one can postulate, as above for tyrosine phosphorylation, that cell adhesion signals transmitted via integrins converge with those triggered by soluble growth factors and oncogenes. The parallels between the platelet and adherent cell systems are extended by the report that binding of fibrinogen to $\alpha_{IIb}\beta_3$ is necessary for the rise in cytoplasmic pH triggered in platelets by epinephrine (Banga et al., 1986).

Therefore, in a variety of cell types, occupation of integrin receptors by their ligands leads to tyrosine phosphorylation and cytoplasmic alkalinization. A reasonable working hypothesis for both these cytoplasmic events (which may indeed be connected) is that integrin receptors synergize with receptors for soluble agonists in stimulating these signals (see below and Figure 4).

Evidence for integrins as costimulatory receptors is also available for T lymphocytes. Ligand engagement of several integrins ($\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_6\beta_1$, and $\alpha_L\beta_2$) can act as a costimulus with cross-linking of the T cell receptor–CD3 receptor on T cells in stimulating cell proliferation (Matsuyama et al., 1989; Shimizu et al., 1990a; Davis et al., 1990; Nojima et al., 1990; van Seventer et al., 1990; Burkly et al., 1991). Furthermore, fibronectin binding to $\alpha_5\beta_1$ has been shown to induce the AP-1 transcription factor necessary for IL-2 transcription (Yamada et al., 1991), and $\alpha_v\beta_3$ has been identified as an accessory costimulator of $\gamma\delta$ T cells for IL-4 production (Roberts et al., 1991). Therefore, extracellular matrix and cell adhesion can have major effects on the activation of T cells, and these effects are mediated via integrins—clearly these are important receptors on lymphocytes, where they play a number of roles (Hemler, 1990; Shimizu and Shaw, 1991).

Integrins can also act as stimulatory receptors for monocytes and neutrophils. Adherence of monocytes to extracellular matrix molecules induces genes encoding inflammatory mediators (Thorens et al., 1987; Sporn et al., 1990). Adherence of neutrophils via β_2 integrins acts as a costimulus with cytokines for induction of the respiratory burst (Nathan et al., 1989), and this adhesion also induces cell motility and Ca^{++} transients in the cytoplasm (Jaconi et al., 1991; Ng-Sikorski et al., 1991).

From all these results, it is clear that integrins can act as true signaling receptors in a variety of cell types. This

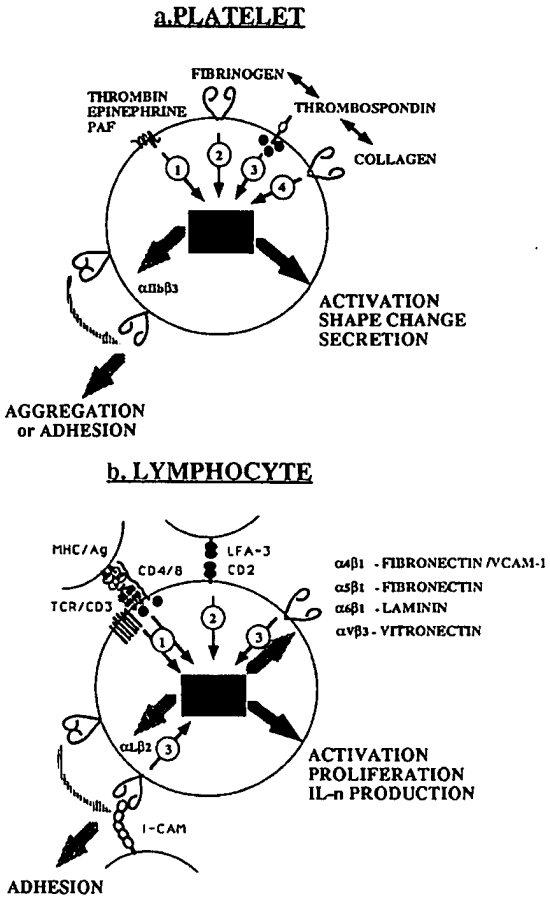


Figure 4. Integrins as Generators and Receivers of Signals

The figure summarizes both data and hypotheses discussed in the text. In each of the three cell types depicted, integrins have been shown to trigger intracellular signals, often in synergy with other receptors. The integrins are depicted as sending signals (thin black arrows) into the second-messenger pathways of the cells (black boxes). These signals converge with those from G protein–coupled agonist receptors or protein tyrosine kinase–coupled receptors. The consequences of cell activation include activation of specific integrins and thus enhanced cell adhesion (large grey arrows) and other responses, such as cell proliferation, secretion, and morphological change (large black arrows).

(a) Platelet. Fibrinogen and collagen each bind to integrins ($\alpha_{IIb}\beta_3$ and $\alpha_2\beta_1$) and can act as agonists. Both ligands also interact with thrombospondin, whose receptor GPIV/CD36 has protein tyrosine kinases (black dots) associated with its cytoplasmic domain. Clustering of the surface receptors is necessary for triggering tyrosine phosphorylation and may facilitate interactions of kinases with their substrates. It is proposed that one consequence of all the signals is activation of $\alpha_{IIb}\beta_3$, leading to strong cell adhesion.

(b) Lymphocyte. Several integrins can act as costimulatory receptors (for signal 3) with the T cell receptor–CD3–CD4,8 receptor complex and its associated kinases (black dots) leading to T cell activation, the consequences of which include activation of integrins for strong cell adhesion to other cells or to extracellular matrix.

(c) Fibroblast. Extracellular matrix molecules, acting in part through integrins, stimulate tyrosine phosphorylation inside the cell, as do growth factor receptors and, in transformed cells, $pp60^{src}$. These signals may converge or synergize to produce cell responses, such as proliferation. Synergy between soluble growth factors and matrix adhesion would give anchorage dependence of growth. Replacement of the need for the extracellular matrix signal by a *src* kinase signal could lead to anchorage independence of transformed cells. The multiple domains of extracellular matrix proteins could cluster receptors to generate combined signals (denoted 1) analogous to those generated by clustering receptors in T cells (see part b).

conclusion offers the potential for explaining results in a wide variety of systems, in which it has been shown that ligand or antibody binding of specific integrins affects gene expression or differentiation of specific cell types. These include induction of specific protease genes in synovial fibroblasts via $\alpha_5\beta_1$ (Werb et al., 1989), inhibition of terminal keratinocyte differentiation by fibronectin acting via a β_1 integrin, probably $\alpha_5\beta_1$ (Adams and Watt, 1989), and modulation of myogenesis (Menko and Boettiger, 1987).

Conclusions and Speculations

From the foregoing discussion, it should be clear that integrins play many roles in many cells. The twenty or so known integrins (see table) offer the possibility of great versatility in cell adhesion, and this versatility is probably further increased by alternative splicing. What has become clear in the last few years is that integrins are not simply adhesion sites on cell surfaces. The activities of many integrins can be radically modulated by cells, and they in turn can modulate cell activities in ways that extend far beyond adhesion.

The view of integrins as two-way signaling molecules is summarized in Figure 3. A particularly important feature of many (perhaps all) integrins is that they undergo activation. It is commonly the case in an adhesion process that integrins provide the strong adhesion but only after activation by other stimuli, which can include soluble mediators (hormones, cytokines, etc.) and/or solid-phase reactants (extracellular matrix or other cells). The specificity of the overall adhesion event lies in the coupling of activation of the final adhesion receptor, often an integrin that is not intrinsically highly specific, to a cascade of signals triggered by specific and/or local events. We discussed four examples: $\alpha_{IIb}\beta_3$ in platelets, $\alpha_L\beta_2$ on lymphocytes, β_2 integrins on leukocytes, and β_1 integrins in lymphocytes. In each case, it is an eminently reasonable model that the adhesion via the integrin is activated at the appropriate time and place by input from more specific signals (respectively, thrombogenic agonists, antigen, selectins plus cytokines, and T cell activation, in the four cases mentioned).

Of equal importance with activation of integrins is their inactivation. It is crucially important that cells should not attach at the wrong times and places. Platelets and leukocytes offer two prime examples in which inappropriate adhesion leads to thrombosis and inflammation, respectively. Thus, the constitutively inactive state of the integrins

on these cells is vital. Similarly, attached cells need on occasion to detach (e.g., during cell migration or mitosis). Less is known about these inactivation events, but hints are beginning to appear that they may be regulated by phosphorylation.

Apart from modulation of affinity, the second major property of integrins reviewed here is their role as signaling receptors. Recent data have uncovered striking parallels in a number of systems (see Figure 3 and earlier discussion). Figure 4 brings together data on several cell systems to allow some parallels to be drawn and some speculations offered. In all three cell types depicted, there is evidence that integrins can act as signaling receptors. In most cases they act as coreceptors with other more traditional receptors, such as the G protein–coupled receptors for platelet agonists or the protein kinase receptors for growth factors in fibroblasts. The signaling events attributed to integrins parallel and synergize with those due to the soluble ligands. This view of integrins as coreceptors fits equally well the results in T lymphocytes, where the antigen-specific T cell receptor (T cell receptor–CD3–CD4,8 complex) is well known to be costimulatory with CD2. Recent data show that various integrins can also serve as coreceptors with the T cell receptor–CD3–CD4 complex.

Thus, in each of these cell types, one can propose a set of signals converging on a common set of regulatory circuits inside the cells (the black boxes in Figure 4). The details of these circuits are unclear but, in each case, appear to include activation of phospholipases, phosphatidylinositol turnover, cytoplasmic alkalinization, elevation of intracellular Ca^{++}, and activation of protein kinase C (All three cell types in Figure 4 as well as leukocytes can be activated by phorbol esters.) In each case, ligand occupancy of integrins can trigger signals that feed into this circuitry and parallel and/or complement the signals from the other receptors.

How do integrins signal? As discussed earlier, there is evidence for activation of protein tyrosine kinases as a consequence of cross-linking of integrins by ligands (fibrinogen or fibronectin) or antibodies. The nature of the kinases involved is unknown, although in platelets, kinases of the *src* family are associated with another adhesion receptor, GPIV/CD36 (Huang et al., 1991; Shattil and Brugge, 1991). Thrombospondin, the ligand of GPIV/CD36 is known to cross-link with fibrinogen and collagen, both ligands for integrins, at the surfaces of aggregating platelets. Thus, it is reasonable to propose that cross-linking of these adhesion receptors brings them together, clustering protein tyrosine kinases in a submembranous patch where they become activated and/or react with their substrates.

There is striking similarity of this model with the current model of T cell activation via the T cell receptor–CD3–CD4 complex (Rudd, 1990; Klausner and Samelson, 1991; Shaw and Thomas, 1991). Here it is thought that protein tyrosine kinases of the *src* family associated with the short cytoplasmic domains of the T cell receptor and CD4/8 are brought together with CD3 by interaction with the complex of antigen and major histocompatibility complex molecule on an antigen-presenting cell, leading to subsequent signaling events. I should stress here that there is as yet no published evidence for tyrosine kinases associated with the short cytoplasmic domains of integrins, but it would certainly be worthwhile looking for them, given the evidence for tyrosine phosphorylation triggered by integrins in several cell types. Also of interest is the 50 kd integrin-associated protein that appears to be involved in integrin-mediated activation of phagocytosis by leukocytes (Brown et al., 1990).

A final speculation is stimulated by the parallels with the T lymphocyte system. It concerns the nature of extracellular matrix proteins with their modular structure and multiple sites for interactions with cell surface receptors, including but not limited to integrins. It has been a puzzling question as to why these molecules have so many distinct and different sites for interacting with cells. Could it be that they are designed to cross-link several different surface receptors together in the plane of the membrane, thus inducing an organized patch of submembrane associates that could interact to generate cytoskeletal structures and/or to trigger signaling events?

The latter proposal is schematized for fibroblasts in Figure 4c, in which an extracellular matrix molecule (e.g., fibronectin) is depicted to interact both with an integrin and with a proteoglycan. The two receptors together generate a signal in a fashion analogous with the signal(s) generated on clustering of T cell receptor–CD3–CD4 in T cells (Figure 4b). One could view the extracellular matrix molecule as the equivalent of the antigen-presenting cell. Evidence exists that segments of fibronectin containing multiple domains trigger greater tyrosine phosphorylation than do simpler cell-binding domains (Guan et al., 1991). Similarly, larger domains promote more cytoskeletal organization (Obara et al., 1988; Woods et al., 1988). Work in progress on various matrix molecules will test this sort of model in detail.

The parallels among the different systems that I have stressed here are undoubtedly more complex than the present discussion implies. Equally, each system has individual features that I have underemphasized, in order to make some generalizations and propose some testable hypotheses concerning functions of integrins in cells. When the integrin receptor family was first recognized several years ago, some people questioned whether integrins were "true" receptors or whether they were "simply" involved in adhesion. The results in the past few years have demonstrated clearly that integrins are indeed receptors, in the sense of transmitting signals both into and out of cells. Furthermore, it has become clear that there is no such thing as simple adhesion. Rather there is a versatile and complex array of interactions, modulations, and signaling events in which integrins play a central role.

Acknowledgments

I would like to thank all my colleagues in the integrin field for sending me numerous reprints and preprints and for stimulating discussions, and Colleen Leslie for her tireless assistance in preparing the manuscript. I gratefully acknowledge support by the Howard Hughes Medical Institute and the National Cancer Institute.

References

Adams, J. C., and Watt, F. M. (1989). Fibronectin inhibits the terminal differentiation of human keratinocytes. Nature *340*, 307–309.

Adams, J. C., and Watt, F. M. (1990). Changes in keratinocyte adhesion during terminal differentiation: reduction in fibronectin binding precedes $\alpha_5\beta_1$ integrin loss from the cell surface. Cell *63*, 425–435.

Albelda, S. M., and Buck, C. A. (1990). Integrins and other cell adhesion molecules. FASEB J. *4*, 2868–2880.

Altieri, D. C., and Edgington, T. S. (1988). A monoclonal antibody reacting with distinct adhesion molecules defines a transition in the functional state of the receptor CD11b/CD18 (Mac-1). J. Immunol. *141*, 2656–2660.

Altruda, F., Cervella, P., Tarone, G., Botta, C., Balzac, F., Stefanuto, G., and Silengo, L. (1990). A human integrin β_1 subunit with a unique cytoplasmic domain generated by alternative mRNA processing. Gene *95*, 261–266.

Andrieux, A., Rabiet, M. J., Chapel, A., Concord, E., and Marguerie, G. (1991). A highly conserved sequence of the Arg-Gly-Asp-binding domain of the integrin β_3 subunit is sensitive to stimulation. J. Biol. Chem. *266*, 14202–14207.

Arnaout, M. A. (1990). Structure and function of the leukocyte adhesion molecules CD11/CD18. Blood *75*, 1037–1050.

Arnaout, M. A., Dana, N., Gupta, S. K., Tenen, D. G., and Fathallah, D.M. (1990). Point mutations impairing cell surface expression of the common β subunit (CD18) in a patient with leukocyte adhesion molecule (Leu-CAM) deficiency. J. Clin. Invest. *85*, 977–981.

Banga, H. S., Simons, E. R., Brass, L. F., and Rittenhouse, S. E. (1986). Activation of phospholipases A and C in human platelets exposed to epinephrine: role of glycoproteins IIb/IIIa and dual role of epinephrine. Proc. Natl. Acad. Sci. USA *83*, 9197–9201.

Bevilacqua, M., Butcher, E., Furie, B., Furie, B., Gallatin, M., Gimbrone, M., Harlan, J., Kishimoto, K., Lasky, L., McEver, R., Paulson, J., Rosen, S., Seed, B., Siegelman, M., Springer, T., Stoolman, L., Tedder, T., Varki, A., Wagner, D., Weissman, I., and Zimmerman, G. (1991). Selectins: a family of adhesion receptors. Cell *67*, 233.

Bodary, S. C., Lipari, T., Muir, C., Napier, M., Pitti, R., and McLean, J. S. (1991). Deletion of the cytoplasmic and transmembrane domains of GPIIbIIIa results in a functional receptor. J. Cell Biol. *115*, 289a.

Bossy, B., Bossy-Wetzel, E., and Reichardt, L. F. (1991). Characterization of the integrin α_8 subunit: a new integrin β_1–associated subunit, which is prominently expressed on axons and on cells in contact with basal laminae in chick embryos. EMBO J. *10*, 2375–2385.

Bray, P. F., Leung, C. S. I., and Shuman, M. A. (1990). Human platelets and megakaryocytes contain alternatively spliced glycoprotein IIb mRNAs. J. Biol. Chem. *265*, 9587–9590.

Brown, E., Hooper, L., Ho, T., and Gresham, H. (1990). Integrin-associated protein: a 50-kD plasma membrane antigen physically and functionally associated with integrins. J. Cell Biol. *111*, 2785–2794.

Brown, N. H., King, D. L., Wilcox, M., and Kafatos, F. C. (1989). Developmentally regulated alternative splicing of Drosophila integrin PS2 α transcripts. Cell *59*, 185–195.

Burkly, L. C., Jakubowski, A., Newman, B. M., Rosa, M. D., Chi-Rosso, G., and Lobb, R. (1991). Signaling by vascular cell adhesion molecule-1 (VCAM-1) through VLA-4 promotes CD-3 dependent T cell proliferation. Eur. J. Immunol. *21*, 2871–2875.

Burridge, K., Fath, K., Kelly, T., Nuckolls, G., and Turner, C. (1988). Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton. Annu. Rev. Cell Biol. *4*, 487–525.

Busk, M., Pytela, R., and Sheppard, D. (1992). Characterization of the integrin $\alpha_v\beta_6$ as a fibronectin-binding protein. J. Biol. Chem., in press.

Butcher, E. C. (1991). Leukocyte–endothelial cell recognition: three (or more) steps to specificity and diversity. Cell *67*, 1033–1036.

Buyon, J. P., Slade, S. G., Reibman, J., Abramson, S. B., Philips, M. R., Weissmann, G., and Winchester, R. (1990). Constitutive and induced phosphorylation of the α- and β-chains of the CD11/CD18 leukocyte integrin family. Relationship to adhesion-dependent functions. J. Immunol. *144*, 191–197.

Calvete, J. J., Henschen, A., and Gonzalez-Rodriguez, J. (1991). Assignment of disulphide bonds in human platelet GPIIIa. A disulphide pattern for the beta-subunits of the integrin family. Biochem. J. *274*, 63–71.

Calvete, J. J., Henschen, A., and Gonzalez-Rodriguez, J. (1989). Complete localization of the intrachain disulphide bonds and the N-glycosylation points in the α-subunit of human platelet glycoprotein IIb. Biochem. J. *261*, 561–568.

Carlos, T. M., and Harlan, J. M. (1990). Membrane proteins involved in phagocyte adherence to endothelium. Immunol. Rev. *114*, 5–24.

Carrell, N. A., Fitzgerald, L. A., Steiner, B., Erickson, H. P., and Phillips, D. R. (1985). Structure of human platelet membrane glycoproteins IIb and IIIa as determined by electron microscopy. J. Biol. Chem. *260*, 1743–1749.

Carter, W. G., Ryan, M. C., and Gahr, P. J. (1991). Epiligrin, a new cell adhesion ligand for integrin $\alpha_3\beta_1$ in epithelial basement membranes. Cell *65*, 599–610.

Chan, B. M., Wong, J. G., Rao, A., and Hemler, M. E. (1991). T cell receptor–dependent, antigen-specific stimulation of a murine T cell clone induces a transient, VLA protein–mediated binding to extracellular matrix. J. Immunol. *147*, 398–404.

Chan, B. M. C., Elices, M. J., Murphy, E., and Hemler, M. E. (1992a). Adhesion to VCAM-1 and fibronectin: comparison of $\alpha_4\beta_1$ (VLA-4) and $\alpha_4\beta_7$ on the human B cell line JY. J. Biol. Chem., in press.

Chan, B. M. C., Kassner, P. D., Schiro, J. A., Byers, R., Kupper, T. S., and Hemler, M. E. (1992b). Distinct cellular functions mediated by different VLA integrin α subunit cytoplasmic domains. Cell *68*, 1051–1060.

Chatila, T. A., Geha, R. S., and Arnaout, M. A. (1989). Constitutive and stimulus-induced phosphorylation of CD11/CD18 leukocyte adhesion molecules. J. Cell Biol. *109*, 3435–3444.

Coller, B. S. (1985). A new murine monoclonal antibody reports an activation-dependent change in the conformation and/or microenvironment of the platelet glycoprotein IIb/IIIa complex. J. Clin. Invest. *76*, 101–108.

Conforti, G., Zanetti, A., Pasquali-Ronchetti, I., Quaglino, D., Jr., Neyroz, P., and Dejana, E. (1990). Modulation of vitronectin receptor binding by membrane lipid composition. J. Biol. Chem. *265*, 4011–4019.

Cooper, H. M., Tamura, R. N., and Quaranta, V. (1991). The major laminin receptor of mouse embryonic stem cells is a novel isoform of the $\alpha_6\beta_1$ integrin. J. Cell Biol. *115*, 843–850.

Corbi, A. L., Garcia-Aguilar, J., and Springer, T. A. (1990). Genomic structure of an integrin α subunit, the leukocyte p150,95 molecule. J. Biol. Chem. *265*, 2782–2788.

D'Souza, S. E., Ginsberg, M. H., Burke, T. A., and Plow, E. F. (1990). The ligand binding site of the platelet integrin receptor GPIIb–IIIa is proximal to the second calcium binding domain of its α subunit. J. Biol. Chem. *265*, 3440–3446.

D'Souza, S. E., Ginsberg, M. H., Burke, T. A., Lam, S. C.-T., and Plow, E. F. (1988). Localization of an Arg-Gly-Asp recognition site within an integrin adhesion receptor. Science *242*, 91–93.

Dahl, S. C., and Grabel, L. B. (1989). Integrin phosphorylation is modulated during the differentiation of F9 teratocarcinoma stem cells. J Cell Biol. *108*, 183–190.

Dana, N., Fathallah, D. M., and Arnaout, M. A. (1991). Expression of a soluble and functional form of the human β_2 integrin CD11b/CD18. Proc. Natl. Acad. Sci. USA *88*, 3106–3110.

Davis, L. S., Oppenheimer-Marks, N., Bednarczyk, J. L., McIntyre, B. W., and Lipsky, P. E. (1990). Fibronectin promotes proliferation of naive and memory T cells by signaling through both the VLA-4 and VLA-5 integrin molecules. J. Immunol. *145*, 785–793.

Diamond, M. S., Staunton, D. E., Marlin, S. D., and Springer, T. A. (1991). Binding of the integrin Mac-1 (CD11b/CD18) to the third immunoglobulin-like domain of ICAM-1 (CD54) and its regulation by glycosylation. Cell *65*, 961–971.

Du, X. P., Plow, E. F., Frelinger, A. L., III, O'Toole, T. E., Loftus, J. C., and Ginsberg, M. H. (1991). Ligands "activate" integrin $\alpha_{IIb}\beta_3$ (platelet GPIIb–IIIa). Cell *65*, 409–416.

Duband, J. L., Rocher, S., Chen, W. T., Yamada, K. M., and Thiery, J.P.

(1986). Cell adhesion and migration in the early vertebrate embryo: location and possible role of the putative fibronectin receptor complex. J. Cell Biol. *102*, 160–178.

Dustin, M. L., and Springer, T. A. (1991). Role of lymphocyte adhesion receptors in transient interactions and cell locomotion. Annu. Rev. Immunol. *9*, 2–66.

Elices, M. J., and Hemler, M. E. (1989). The human integrin VLA-2 is a collagen receptor on some cells and a collagen/laminin receptor on others. Proc. Natl. Acad. Sci. USA *86*, 9906–9910.

Elices, M. J., Urry, L. A., and Hemler, M. E. (1991). Receptor functions for the integrin VLA-3: fibronectin, collagen, and laminin binding are differentially influenced by Arg-Gly-Asp peptide and by divalent cations. J. Cell Biol. *112*, 169–181.

Erle, D. J., Ruegg, C., Sheppard, D., and Pytela, R. (1991). Complete amino acid sequence of an integrin beta subunit (β_7) identified in leukocytes. J. Biol. Chem. *266*, 11009–11016.

Ferguson, T. A., Mizutani, H., and Kupper, T. S. (1991). Two integrin-binding peptides abrogate T cell–mediated immune responses *in vivo*. *Proc. Natl. Acad. Sci. USA 88*, 8072–8076.

Ferreira, O. C., Jr., Valinsky, J. E., Sheridan, K., Wayner, E. A., Bianco, C., and Garcia-Pardo, A. (1991). Phorbol ester–induced differentiation of U937 cells enhances attachment to fibronectin and distinctly modulates the $\alpha_5\beta_1$ and $\alpha_4\beta_1$ fibronectin receptors. Exp. Cell Res. *193*, 20–26.

Ferrell, J. E., and Martin, G. S. (1989). Tyrosine-specific protein phosphorylation is regulated by glycoprotein IIb–IIIa in platelets. Proc. Natl. Acad. Sci. USA *86*, 2234–2238.

Frelinger, A. L., III, Lam, S. C.-T., Plow, E. F., Smith, M. A., Loftus, J. C., and Ginsberg, M. H. (1988). Occupancy of an adhesive glycoprotein receptor modulates expression of an antigenic site involved in cell adhesion. J. Biol. Chem. *263*, 12397–12402.

Frelinger, A. L., III, Cohen, I., Plow, E. F., Smith, M. A., Roberts, J., Lam, S. C.-T., and Ginsberg, M. H. (1990). Selective inhibition of integrin function by antibodies specific for ligand-occupied receptor conformers. J. Biol. Chem. *104*, 1655–1663.

Frelinger, A. L., III, Du, X. P., Plow, E. F., and Ginsberg, M. H. (1991). Monoclonal antibodies to ligand-occupied conformers of integrin $\alpha_{IIb}\beta_3$ (glycoprotein IIb–IIIa) alter receptor affinity, specificity, and function. J. Biol. Chem. *266*, 17106–17111.

Gailit, J., and Ruoslahti, E. (1988). Regulation of the fibronectin receptor affinity by divalent cations. J. Biol. Chem. *263*, 12927–12933.

Golden, A., Brugge, J. S., and Shattil, S. J. (1990). Role of platelet membrane glycoprotein IIb–IIIa in agonist-induced tyrosine phosphorylation of platelet proteins. J. Cell Biol. *111*, 3117–3127.

Guan, J.-L., Trevithick, J. E., and Hynes, R. O. (1991). Fibronectin/integrin interaction induces tyrosine phosphorylation of a 120 kDa protein. Cell Regul. *2*, 951–964.

Gulino, D., Ryckewaert, J. J., Andrieux, A., Rabiet, M. J., and Marguerie, G. (1990). Identification of a monoclonal antibody against platelet GPIIb that interacts with a calcium-binding site and induces aggregation. J. Biol. Chem. *265*, 9575–9581.

Hall, D. E., Reichardt, L. F., Crowley, E., Holley, B., Moezzi, H., Sonnenberg, A., and Damsky, C. H. (1990). The α_6/β_1 integrin heterodimers mediate cell attachment to distinct sites on laminin. J. Cell Biol. *110*, 2175–2184.

Hayashi, Y., Haimovich, B., Reszka, A., Boettiger, D., and Horwitz, A. (1990). Expression and function of chicken integrin β_1 subunit and its cytoplasmic domain mutants in mouse NIH 3T3 cells. J. Cell Biol. *110*, 175–184.

Heidenreich, R., Elsman, R., Surrey, S., Delgrosso, K., Bennett, J. S., Schwartz, E., and Poncz, M. (1990). Organization of the gene for platelet glycoprotein IIb. Biochemistry *29*, 1232–1244.

Hemler, M. E. (1990). VLA proteins in the integrin family: structures, functions and their role on leukocytes. Annu. Rev. Immunol. *8*, 365–400.

Hermanowski-Vosatka, A., Van Strijp, J. A. G., Swiggard, W. J., and Wright, S. D. (1992). Integrin modulating factor-1: a lipid that alters the function of leukocyte integrins. Cell *68*, 341–352.

Hibbs, M. L., Xu, H., Stacker, S. A., and Springer, T. A. (1991a). Regulation of adhesion of ICAM-1 by the cytoplasmic domain of LFA-1 integrin β subunit. Science *251*, 1611–1613.

Hibbs, M. L., Jakes, S., Stacker, S. A., Wallace, R. W., and Springer, T. A. (1991b). The cytoplasmic domain of the integrin lymphocyte function–associated antigen 1 β subunit: sites required for binding to intercellular adhesion molecule 1 and the phorbol ester–stimulated phosphorylation site. J. Exp. Med. *174*, 1227–1238.

Hillery, C. A., Smyth, S. S., and Parise, L. V. (1991). Phosphorylation of human platelet glycoprotein IIIa (GPIIIa). Dissociation from fibrinogen receptor activation and phosphorylation of GPIIIa *in vitro*. J. Biol. Chem. *266*, 14663–14669.

Hogervorst, F., Kuikman, I., van Kessel, A. G., and Sonnenberg, A. (1991). Molecular cloning of the human α_6 integrin subunit. Alternative splicing of α_6 mRNA and chromosomal localization of the α_6 and β_4 genes. Eur. J. Biochem. *199*, 425–433.

Hogervorst, F., Kuikman, I., von dem Borne, A. E. G. K., and Sonnenberg, A. (1990). Cloning and sequence analysis of beta-4 cDNA: an integrin subunit that contains a unique 118 kd cytoplasmic domain. EMBO J. *9*, 765–770.

Horvath, A. R., Elmore, M. A., and Kellie, S. (1990). Differential tyrosine-specific phosphorylation of integrin in Rous sarcoma virus transformed cells with differing transformed phenotypes. Oncogene *5*, 1349–1357.

Horwitz, A., Duggan, E., Buck, C., Beckerle, M. C., and Burridge, K. (1986). Interaction of plasma membrane fibronectin receptor with talin—a transmembrane linkage. Nature *320*, 531–533.

Huang, M. M., Bolen, J. B., Barnwell, J. W., Shattil, S. J., and Brugge, J. S. (1991). Membrane glycoprotein IV (CD36) is physically associated with Fyn, Lyn and Yes protein tyrosine kinases in human platelets. Proc. Natl. Acad. Sci. USA *88*, 7844–7848.

Hynes, R. O. (1987). Integrins: a family of cell surface receptors. Cell *48*, 549–554.

Hynes, R. O., and Lander, A. D. (1992). Contact and adhesive specificities in the associations, migrations, and targeting of cells and axons. Cell *68*, 303–322.

Hynes, R. O., Marcantonio, E. E., Stepp, M. A., Urry, L. A., and Yee, G. (1989). Integrin heterodimer and receptor complexity in avian and mammalian cells. J. Cell Biol. *109*, 409–420.

Ingber, D. E., Prusty, D., Frangioni, J. V., Cragoe, E. J., Jr., Lechene, C., and Schwartz, M. A. (1990). Control of intracellular pH and growth by fibronectin in capillary endothelial cells. J. Cell Biol. *110*, 1803–1811.

Jaconi, M. E. E., Theler, J. M., Schlegel, W., Appel, R. D., Wright, S. D., and Lew, P. D. (1991). Multiple elevations of cytosolic-free Ca^{2+} in human neutrophils: initiation by adherence receptors of the integrin family. J. Cell Biol. *112*, 1249–1257.

Kajiji, S., Tamura, R. N., and Quaranta, V. (1989). A novel integrin ($\alpha_E\beta_4$) from human epithelial cells suggests a fourth family of integrin adhesion receptors. EMBO J. *8*, 673–680.

Kanner, S. B., Reynolds, A. B., Vines, R. R., and Parsons, J. T. (1990). Monoclonal antibodies to individual tyrosine-phosphorylated protein substrates of oncogene-encoded tyrosine kinases. Proc. Natl. Acad. Sci. USA *87*, 3328–3332.

Keizer, G. D., Visser, W., Vliem, M., and Figdor, C. G. (1988). A monoclonal antibody (NKI-L16) directed against a unique epitope of the α-chain of human leukocyte function–associated antigen 1 induces homotypic cell–cell interactions. J. Immunol. *140*, 1393–1400.

Kelly, T., Molony, L., and Burridge, K. (1987). Purification of two smooth muscle glycoproteins related to integrin. Distribution in cultured chicken embryo fibroblasts. J. Biol. Chem. *262*, 17189–17199.

Kieffer, N., and Phillips, D. R. (1990). Platelet membrane glycoproteins: functions in cellular interactions. Annu. Rev. Cell Biol. *6*, 329–357.

Kieffer, N., Fitzgerald, L. A., Wolf, D., Cheresh, D. A., and Phillips, D. R. (1991). Adhesive properties of the β_3 integrins: comparison of GPIIb–IIIa and the vitronectin receptor individually expressed in human melanoma cells. J. Cell Biol. *113*, 451–461.

Kirchhofer, D., Gailit, J., Ruoslahti, E., Grzesiak, J., and Pierschbacher, M. D. (1990a). Cation-dependent changes in the binding specificity of the platelet receptor GPIIb/IIIa. J. Biol. Chem. *265*, 18525–18530.

Kirchhofer, D., Languino, L. R., Ruoslahti, E., and Pierschbacher, M. D. (1990b). $\alpha_2\beta_1$ integrins from different cell types show different binding specificities. J. Biol. Chem. *265*, 615–618.

Kirchhofer, D., Grzesiak, J., and Pierschbacher, M. D. (1991). Calcium as a potential physiological regulator of integrin-mediated cell adhesion. J. Biol. Chem. *266*, 4471–4477.

Kishimoto, T. K., O'Connor, K., and Springer, T. A. (1989). Leukocyte adhesion deficiency. Aberrant splicing of a conserved integrin sequence causes a moderate deficiency phenotype. J. Biol. Chem. *264*, 3588–3595.

Klausner, R. D., and Samelson, L. E. (1991). T cell antigen receptor activation pathways: the tyrosine kinase connection. Cell *64*, 875–878.

Koch, C. A., Anderson, D., Moran, M. J., Ellis, C., and Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. Science *252*, 668–674.

Kornberg, L. J., Earp, H. S., Turner, C. E., Prockop, C., and Juliano, R. L. (1991). Signal transduction by integrins: increased protein tyrosine phosphorylation caused by clustering of β_1 integrins. Proc. Natl. Acad. Sci. USA *88*, 8392–8396.

Kouns, W. C., Wall, C. D., White, M. M., Fox, C. F., and Jennings, L. K. (1990). A conformation-dependent epitope of human platelet glycoprotein IIIa. J. Biol. Chem. *265*, 20594–20601.

Kramer, R. H., Vu, M. P., Cheng, Y. F., Ramos, D. M., Timpl, R., and Waleh, N. (1991). Laminin-binding integrin $\alpha_7\beta_1$: functional characterization and expression in normal and malignant melanocytes. Cell Regul. *2*, 805–817.

Kurpakus, M. A., Quaranta, V., and Jones, J. C. R. (1991). Surface relocation of alpha$_6$beta$_4$ integrins and assembly of hemidesmosomes in an in vitro model of wound healing. J. Cell Biol. *115*, 1737–1750.

Larson, R. S., and Springer, T. A. (1990). Structure and function of leukocyte integrins. Immunol. Rev. *114*, 181–217.

Larson, R. S., Hibbs, M. L., and Springer, T. A. (1991). The leukocyte integrin LFA-1 reconstituted by cDNA transfection in a nonhematopoietic cell line is functionally active and not transiently regulated. Cell Regul. *1*, 359–367.

Lasky, L. A., and Rosen, S. D. (1992). The selectins: carbohydrate-binding adhesion molecules of the immune system. In Inflammation: Basic Principles and Clinical Correlates, J. Gallin, I. Goldstein, and R. Snyderman, eds. (New York: Raven Press), in press.

Lawrence, M. B., and Springer, T. A. (1991). Leukocytes roll on a selectin at physiologic flow rates: distinction from and prerequisite for adhesion through integrins. Cell *65*, 859–873.

Loftus, J. C., O'Toole, T. E., Plow, E. F., Glass, A., Frelinger, A. L., III, and Ginsberg, M. H. (1990). A β_3 integrin mutation abolishes ligand binding and alters divalent cation–dependent conformation. Science *249*, 915–918.

Loike, J. D., Sodeik, B., Cao, L., Leucona, S., Weitz, J. I., Detmers, P. A., Wright, S. D., and Silverstein, S. C. (1991). CD11c/CD18 on neutrophils recognizes a domain at the N terminus of the Aα chain of fibrinogen. Proc. Natl. Acad. Sci. USA *88*, 1044–1048.

Lotz, M. M., Korzelius, C. A., and Mercurio, A. M. (1991). Human colon carcinoma cells use multiple receptors to adhere to laminin: involvement of $\alpha_6\beta_4$ and $\alpha_2\beta_1$ integrins. Cell Regul. *1*, 249–257.

Maher, P. A., Pasquale, E. B., Wang, J. Y. J., and Singer, S. J. (1985). Phosphotyrosine-containing proteins are concentrated in focal adhesions and intercellular junctions in normal cells. Proc. Natl. Acad. Sci. USA *82*, 6576–6580.

Manning, D. R., and Brass, L. F. (1991). The role of GTP-binding proteins in platelet activation. Thromb. Haemost. *66*, 393–399.

Marcantonio, E. E., Guan, J. L., Trevithick, J. E., and Hynes, R. O. (1990). Mapping of the functional determinants of the integrin β_1 cytoplasmic domain by site-directed mutagenesis. Cell Regul. *1*, 597–604.

Matsuyama, Y., Yamada, A., Kay, J., Yamada, K. M., Akiyama, S. K., Schlossman, S. F., and Morimoto, C. (1989). Activation of CD4 cells by fibronectin and anti-CD3 antibody. J. Exp. Med. *170*, 1133–1148.

McLean, J. W., Vestal, D. J., Cheresh, D. A., and Bodary, S. C. (1990). cDNA sequence of the human integrin beta 5 subunit. J. Biol. Chem. *265*, 17126–17131.

Menko, A. S., and Boettiger, D. (1987). Occupation of the extracellular matrix receptor, integrin, is a control point for myogenic differentiation. Cell *51*, 51–57.

Moyle, M., Napier, M. A., and McLean, J. W. (1991). Cloning and expression of a divergent integrin subunit β_8. J. Biol. Chem. *266*, 19650–19658.

Nathan, C., Srimal, S., Farber, C., Sanchez, E., Kabbash, L., Asch, A., Gailit, J., and Wright, S. D. (1989). Cytokine-induced respiratory burst of human neutrophils: dependence on extracellular matrix proteins and CD11/CD18 integrins. J. Cell Biol. *109*, 1341–1349.

Nermut, M. V., Green, N. M., Eason, P., Yamada, S. S., and Yamada, K.M. (1988). Electron microscopy and structural model of human fibronectin receptor. EMBO J. *7*, 4093–4099.

Neugebauer, K. M., and Reichardt, L. F. (1991). Cell-surface regulation of β_1–integrin activity on developing retinal neurons. Nature *350*, 68–71.

Ng-Sikorski, J., Andersson, R., Patarroyo, M., and Andersson, T. (1991). Calcium signaling capacity of the CD11b/CD18 integrin on human neutrophils. Exp. Cell Res. *195*, 504–508.

Nojima, Y., Humphries, M. J., Mould, A. P., Komoriya, A., Yamada, K. M., Schlossman, S. F., and Morimoto, C. (1990). VLA-4 mediates CD3-dependent CD4$^+$ T cell activation via the CS1 alternatively spliced domain of fibronectin. J. Exp. Med. *172*, 1185–1192.

O'Toole, T. E., Loftus, J. C., Du, X. P., Glass, A. A., Ruggeri, Z. M., Shattil, S. J., Plow, E.F., and Ginsberg, M. H. (1991a). Affinity modulation of the $\alpha_{IIb}\beta_3$ integrin (platelet GPIIb–IIIa) is an intrinsic property of the receptor. Cell Regul. *1*, 883–893.

O'Toole, T. E., Mandelman, D., Forsyth, J., Shattil, S. J., Plow, E.F., and Ginsberg, M. H. (1991b). Modulation of the affinity of integrin $\alpha_{IIb}\beta_3$ (GPIIB–IIIa) by the cytoplasmic domain of α_{IIb}. Science *254*, 845–847.

Obara, M., Kang, M. S., and Yamada, K. M. (1988). Site-directed mutagenesis of the cell-binding domain of human fibronectin: separable, synergistic sites mediate adhesive function. Cell *53*, 649–657.

Osborn, L. (1990). Leukocyte adhesion to endothelium in inflammation. Cell *62*, 3–6.

Parise, L. V., Helgerson, S. L., Steiner, B., Nannizzi, L., and Phillips, D. R. (1987). Synthetic peptides derived from fibrinogen and fibronectin change the conformation of purified platelet glycoprotein IIb–IIIa. J. Biol. Chem. *262*, 12597–12602.

Parker, C. M., Cepek, K., Russell, G. J., Shaw, S. K., Posnett, D., Schwarting, R., and Brenner, M. B. (1992). A family of β_7 integrins on human mucosal lymphocytes. Proc. Natl. Acad. Sci. USA *89*, 1924–1928.

Pasquale, E. B., Maher, P. A., and Singer, S. J. (1988). Comparative study of the tyrosine phosphorylation of proteins in Swiss 3T3 fibroblasts stimulated by a variety of mitogenic agents. J. Cell. Physiol. *137*, 146–156.

Phillips, D. R., Charo, I. F., and Scarborough, R. M. (1991). GPIIb–IIIa: the responsive integrin. Cell *65*, 359–362.

Ramaswamy, H., and Hemler, M. E. (1990). Cloning, primary structure and properties of a novel human integrin β subunit. EMBO J. *9*, 1561–1568.

Rees-Jones, R. W., and Taylor, S. I. (1985). An endogenous substrate for the insulin receptor–associated tyrosine kinase. J. Biol. Chem. *260*, 4461–4467.

Roberts, K., Yokoyama, W. M., Kehn, P. J., and Shevach, E. M. (1991). The vitronectin receptor serves as an accessory molecule for the activation of a subset of gamma/delta T cells. J. Exp. Med. *173*, 231–240.

Rudd, C. E. (1990). CD4, CD8 and the TCR–CD3 complex: a novel class of protein–tyrosine kinase receptor. Immunol. Today *11*, 400–406.

Rüegg, C., Postigo, A. A., Sikorski, E. E., Butcher, E. C., Pytela, R., and

Erle, D. J. (1992). Role of $\alpha_4\beta_7/\alpha_4\beta_P$ integrin in lymphocyte adherence to fibronectin and VCAM-1 and in homotypic cell clustering. J. Cell Biol., in press.

Ruoslahti E. (1991). Integrins. J. Clin. Invest. *87*, 1–5.

Sadoul, J. L., Peyron, J.-F., Ballotti, R., Debant, A., Fehlmann, M., and Van Obberghen, E. (1985). Identification of a cellular 110,000 Da protein substrate for the insulin receptor kinase. Biochem. J. *227*, 887–892.

Schwartz, B. R., Wayner, E. A., Carlos, T. M., Ochs, H. D., and Harlan, J. M. (1990). Identification of surface proteins mediating adherence of CD11/CD18-deficient lymphoblastoid cells to cultured human endothelium. J. Clin. Invest. *85*, 2019–2022.

Schwartz, M. A., Both, G., and Lechene, C. (1989). Effect of cell spreading on cytoplasmic pH in normal and transformed fibroblasts. Proc. Natl. Acad. Sci. USA *86*, 4525–4529.

Schwartz, M. A., Rupp, E. E., Frangioni, J. V., and Lechene, C. P. (1990). Cytoplasmic pH and anchorage-independent growth induced by v-Ki-*ras*, v-*src* or polyoma middle T. Oncogene *5*, 55–58.

Schwartz, M. A., Ingber, D. E., Lawrence, M., Springer, T. A., and Lechene, C. (1991a). Multiple integrins share the ability to induce elevation of intracellular pH. Exp. Cell Res. *195*, 533–535.

Schwartz, M. A., Lechene, C., and Ingber, D. E. (1991b). Insoluble fibronectin activates the Na/H antiporter by clustering and immobilizing integrin $\alpha_5\beta_1$, independent of cell shape. Proc. Natl. Acad. Sci. USA *88*, 7849–7853.

Shattil, S. J., and Brugge, J. S. (1991). Protein tyrosine phosphorylation and the adhesive functions of platelets. Curr. Opin. Cell Biol. *3*, 869–879.

Shattil, S. J., Hoxie, J. A., Cunningham, M., and Brass, L. F. (1985). Changes in the platelet membrane glycoprotein IIb–IIIa complex during platelet activation. J. Biol. Chem. *260*, 11107–11114.

Shaw, A., and Thomas, M. L. (1991). Coordinate interactions of protein tyrosine kinases and protein tyrosine phosphatases in T-cell receptor-mediated signalling. Curr. Opin. Cell Biol. *3*, 862–868.

Sheppard, D., Rozzo, C., Starr, L., Quaranta, V., Erle, D. J., and Pytela, R. (1990). Complete amino acid sequence of a novel integrin β subunit (β_6) identified in epithelial cells using the polymerase chain reaction. J. Biol. Chem. *265*, 11502–11507.

Shimizu, Y., van Seventer, G. A., Horgan, K. J., and Shaw, S. (1990a). Costimulation of proliferative responses of resting $CD4^+$ T cells by the interaction of VLA-4 and VLA-5 with fibronectin or VLA-6 with laminin. J. Immunol. *145*, 59–67.

Shimizu, Y., van Seventer, G., Horgan, K. J., and Shaw, S. (1990b). Regulated expression and binding of three VLA (β_1) integrin receptors on T cells. Nature *345*, 250–253.

Shimizu, Y., and Shaw S. (1991). Lymphocyte interactions with extracellular matrix. FASEB J. *5*, 2292–2299.

Shimizu, Y., Newman, W., Gopal, T. V., Horgan, K. J., Graber, N., Beal, L. D., van Seventer, G. A., and Shaw, S. (1991). Four molecular pathways of T cell adhesion to endothelial cells: roles of LFA-1, VCAM-1, and ELAM-1 and changes in pathway hierarchy under different activation conditions. J. Cell Biol. *113*, 1203–1212.

Sims, P. J., Ginsberg, M. H., Plow, E. F., and Shattil, S. J. (1991). Effect of platelet activation on the conformation of the plasma membrane glycoprotein IIb–IIIa complex. J. Biol. Chem. *266*, 7345–7352.

Smith, C. W., Kishimoto, T. K., Abbass, O., Hughes, B., Rothlein, R., McIntire L. V., Butcher, E., and Anderson, D. C. (1991). Chemotactic factors regulate lectin adhesion molecule 1 (LECAM-1)-dependent neutrophil adhesion to cytokine-stimulated endothelial cells in vitro. J. Clin. Invest. *87*, 609–618.

Smith, J. W., and Cheresh, D. A. (1988). The Arg-Gly-Asp binding domain of the vitronectin receptor. Photoaffinity cross-linking implicates amino acid residues 61–203 of the β subunit. J. Biol. Chem. *263*, 18726–18731.

Smith, J. W., and Cheresh, D. A. (1990). Integrin $(\alpha_v\beta_3)$–ligand interaction. J. Biol. Chem. *265*, 2168–2172.

Solowska, J., Edelman, J. M., Albelda, S. M., and Buck, C. A. (1991). Cytoplasmic and transmembrane domains of integrin β_1 and β_3 subunits are functionally interchangeable. J. Cell Biol. *114*, 1079–1088.

Solowska, J., Guan, J.-L., Marcantonio, E. E., Trevithick, J. E., Buck, C. A., and Hynes, R. O. (1989). Expression of normal and mutant avian integrin subunits in rodent cells. J. Cell Biol. *109*, 853–861.

Sonnenberg, A., Linders, C. J. T., Modderman, P. W., Damsky, C. H., Aumailley, M., and Timpl, R. (1990). Integrin recognition of different cell-binding fragments of laminin (P1, E3, E8) and evidence that $\alpha_6\beta_1$ but not $\alpha_6\beta_4$ functions as a major receptor for fragment E8. J. Cell Biol. *110*, 2145–2155.

Sonnenberg, A., Calafat, J., Janssen, H., Daams, H., van der Raaij-Helmer, L. M., Falcioni, R., Kennel, S. J., Aplin, J. D., Baker, J., Loizidou, M., and Garrod, D. (1991). Integrin α_6/β_4 complex is located in hemidesmosomes, suggesting a major role in epidermal cell–basement membrane adhesion. J. Cell Biol. *113*, 907–917.

Sporn, S. A., Eierman, D. F., Johnson, C. E., Morris, J., Martin, G., Ladner, M., and Haskill, S. (1990). Monocyte adherence results in selective induction of novel genes sharing homology with mediators of inflammation and tissue repair. J. Immunol. *144*, 4434–4441.

Springer, T. A. (1990a). Adhesion receptors of the immune system. Nature *346*, 425–434.

Springer, T. A. (1990b). The sensation and regulation of interactions with the extracellular environment: the cell biology of lymphocyte adhesion receptors. Annu. Rev. Cell Biol. *6*, 359–402.

Staatz, W. D., Rajpara, S. M., Wayner, E. A., Carter, W. G., and Santoro, S. A. (1989). The membrane glycoprotein Ia–IIa (VLA-2) complex mediates the Mg^{++}-dependent adhesion of platelets to collagen. J. Cell Biol. *108*, 1917–1924.

Staatz, W. D., Fok, K. F., Zutter, M. M., Adams, S. P., Rodriguez, B. A., and Santoro, S.A. (1991). Identification of a tetrapeptide recognition sequence for the $\alpha_2\beta_1$ integrin in collagen. J. Biol. Chem. *266*, 7363–7367.

Staunton, D. E., Dustin, M. L., Erickson, H. P., and Springer, T.A. (1990). The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. Cell *61*, 243–254.

Stepp, M. A., Spurr-Michaud, S., Tisdale, A., Elwell, J., and Gipson, I. K. (1990). $\alpha_6\beta_4$ integrin heterodimer is a component of hemidesmosomes. Proc. Natl. Acad. Sci. USA *87*, 8970–8974.

Suzuki, S., and Naitoh, Y. (1990). Amino acid sequence of a novel integrin β_4 subunit and primary expression of the mRNA in epithelial cells. EMBO J. *9*, 757–763.

Suzuki, S., Huang, Z. S., and Tanihara, H. (1990). Cloning of an integrin β subunit exhibiting high homology with integrin β_3 subunit. Proc. Natl. Acad. Sci. USA *87*, 5354–5358.

Tamura, R. N., Rozzo, C., Starr, L., Chambers, J., Reichardt, L. F., Cooper, H. M., and Quaranta, V. (1990). Epithelial integrin $\alpha_6\beta_4$: complete primary structure of α_6 and variant forms of β_4. J. Cell Biol. *111*, 1593–1604.

Tamura, R. N., Cooper, H. M., Collo, G., and Quaranta, V. (1991). Cell type–specific integrin variants with alternative α chain cytoplasmic domains. Proc. Natl. Acad. Sci. USA *88*, 10183–10187.

Tapley, P., Horwitz, A. F., Buck, C. A., Burridge, K., Duggan, K., Hirst, R., and Rohrschneider, L. (1989). Analysis of the avian fibronectin receptor (integrin) as direct substrate for $pp60^{v\text{-}src}$. Oncogene *4*, 325–333.

Thorens, B., Mermod, J.-J., and Vassalli, P. (1987). Phagocytosis and inflammatory stimuli induce GM-CSF mRNA in macrophages through posttranscriptional regulation. Cell *48*, 671–679.

Tsukita, S., Oishi, K., Akiyama, T., Yamanashi, Y., Yamamoto, T., and Tsukita, S. (1991). Specific proto-oncogenic tyrosine kinases of src family are enriched in cell-to-cell adherens junctions where the level of tyrosine phosphorylation is elevated. J. Cell Biol. *113*, 867–879.

Valmu, L., Autero, M., Siljander, P., Patarroyo, M., and Gahmberg, C. G. (1991). Phosphorylation of the β-subunit of CD11/CD18 integrins by protein kinase C correlates with leukocyte adhesion. Eur. J. Immunol. *21*, 2857–2862.

van Kooyk, Y., van de Wiel-van Kamenade, P., Weder, P., Kuijpers, T. W., and Figdor, C. G. (1989). Enhancement of LFA-1-mediated cell adhesion by triggering through CD2 or CD3 on T lymphocytes. Nature *342*, 811–813.

van Kooyk, Y., Weder, P., Hogervorst, F., Verhoeven, A. J., van Seventer, G., te Velde, A. A., Borst, J., Keizer, G. D., and Figdor, C. G. (1991). Activation of LFA-1 through a calcium-dependent epitope stimulates lymphocyte adhesion. J. Cell Biol. *112*, 345–354.

van Kuppevelt, T. H. M. S. M., Languino, L. R., Gailit, J. O., Suzuki, S., and Ruoslahti, E. (1989). An alternative cytoplasmic domain of the integrin β_3 subunit. Proc. Natl. Acad. Sci. USA *86*, 5415–5418.

van Seventer, G. A., Shimizu, Y., Horgan, K. J., and Shaw, S. (1990). The LFA-1 ligand ICAM-1 provides an important costimulatory signal for T cell receptor–mediated activation of resting T cells. J. Immunol. *144*, 4579–4586.

Vogel, B. E., Tarone, G., Giancotti, F. G., Gailit, J., and Ruoslahti, E. (1990). A novel fibronectin receptor with an unexpected subunit composition ($\alpha_v\beta_1$). J. Biol. Chem. *265*, 5934–5937.

von Andrian, U. H., Chambers, J. D., McEvoy, L. M., Bargatze, R. F., Arfors, K.-E., and Butcher, E. C. (1991). Two-step model of leukocyte–endothelial cell interaction in inflammation: distinct roles for LECAM-1 and leukocyte β_2 integrins in vivo. Proc. Natl. Acad. Sci. USA *88*, 7538–7542.

von der Mark, H., Dürr, J., Sonnenberg, A., von der Mark, K., Deutzmann, R., and Goodman, S. L. (1991). Skeletal myoblasts utilize a novel β_1-series integrin and not α_6 β_1 for binding to the E8 and T8 fragments of laminin. J. Biol. Chem. *266*, 23593–23601.

Wardlaw, A. J., Hibbs, M. L., Stacker, S. A., and Springer, T. A. (1990). Distinct mutations in two patients with leukocyte adhesion deficiency (LAD) and their functional correlates. J. Exp. Med. *172*, 335–345.

Wayner, E. A., Carter, W. G., Piotrowicz, R. S., and Kunicki, T. J. (1988). The function of multiple extracellular matrix receptors in mediating cell adhesion to extracellular matrix: preparation of monoclonal antibodies to the fibronectin receptor that specifially inhibit cell adhesion to fibronectin and react with platelet glycoproteins Ic–IIa. J. Cell Biol. *107*, 1881–1891.

Wayner, E. A., Orlando, R. A., and Cheresh, D. A. (1991). Integrins $\alpha_v\beta_3$ and $\alpha_v\beta_5$ contribute to cell attachment to vitronectin but differentially distribute on the cell surface. J. Cell Biol. *113*, 919–929.

Werb, Z., Tremble, P. M., Behrendtsen, O., Crowley, E., and Damsky, C. H. (1989). Signal transduction through the fibronectin receptor induces collagenase and stromelysin gene expression. J. Cell Biol. *109*, 877–889.

Wilkins, J. A., Stupack, D., Stewart, S., and Caixia, S. (1991). β_1 integrin–mediated lymphocyte adherence to extracellular matrix is enhanced by phorbol ester treatment. Eur. J. Immunol. *21*, 517–522.

Woods, A., Johansson, S., and Hook, M. (1988). Fibronectin fibril formation involves cell interactions with two fibronectin domains. Exp. Cell Res. *177*, 272–283.

Wright, S. D., and Silverstein, S. C. (1982). Tumor-promoting phorbol esters stimulate C3b and C3bi receptor–mediated phagocytosis in cultured human monocytes. J. Exp. Med. *156*, 1149–1164.

Wright, S. D., Rao, P. E., Van Voorhis, W. C., Craigmyl, L. S., Jida, K., Talle, M. A., Westberg, E. F., Goldstein, G., and Silverstein, S. C. (1983). Identification of the C3bi receptor on human monocytes and macrophages by using monoclonal antibodies. Proc. Natl. Acad. Sci. USA *80*, 5699–5703.

Wright, S. D., Licht, M. R., Craigmyl, L. S., and Silverstein, S.C. (1984). Communication between receptors for different ligands on a single cell: ligation of fibronectin receptors induces a reversible alteration in the function of complement receptors on cultured human monocytes. J. Cell Biol. *99*, 336–339.

Yamada, A., Nikaido, T., Nojima, Y., Schlossman, S. F., and Morimoto, C. (1991). Activation of human CD4 T lymphocytes. Interaction of fibronectin with VLA-5 receptor on CD4 cells induces the AP-1 transcription factor. J. Immunol. *146*, 53–56.

Yuan, Q. A., Jiang, W. M., Krissansen, G. W., and Watson, J. D. (1990). Cloning and sequence analysis of a novel β_2-related integrin transcript from T lymphocytes: homology of integrin cysteine-rich repeats to domain III of laminin B chains. Int. Immunol. *2*, 1097–1108.

Yuan, Q., Jiang, W. M., Hollander, D., Leung, E., Watson, J. D., and Krissansen, G. W. (1991). Identity between the novel integrin β_7 subunit and an antigen found highly expressed on intraepithelial lymphocytes in the small intestine. Biochem. Biophys. Res. Comm. *176*, 1443.

Note Added in Proof

The sequence of rat α_7 has been completed (Song, W. K., Wang, W., Foster, R. F., Bielser, D. A., and Kaufman, S. J. (1992). H36-α_7 is a novel integrin alpha chain that is developmentally regulated during skeletal myogenesis. J. Cell Biol., in press.)

U. S. A. 95, 11199–11204
24 Takei, K. *et al.* (1998) Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes. *Cell* 94, 131–141
25 Matsuoka, K. *et al.* (1998) Coat assembly directs v-SNARE concentration into synthetic COPII vesicles. *Mol. Cell* 2, 703–708
26 Faundez, F. *et al.* (1998) A function for the AP3 coat complex in synaptic vesicle formation from endosomes. *Cell* 93, 423–432
27 Owen, D.J. and Evans, P.R. (1998) A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* 282, 1327–1332
28 Blagoveschenskaya, A.D. *et al.* (1999) A complex web of signal-dependent trafficking underlies the triorganellar distribution of P-selectin in neuroendocrine PC12 cells. *J. Cell Biol.* 145, 1419–1433
29 Chen, C.S. *et al.* (1997) Changes in the spectral properties of a plasma membrane lipid analog during the first seconds of endocytosis in living cells. *Biophys. J.* 72, 37–50
30 Mukherjee, S. *et al.* (1999) Endocytic sorting of lipid analogues differing solely in the chemistry of their hydrophobic tails. *J. Cell Biol.* 144, 1271–1284
31 Simons, K. and Ikonen, E. (1997) Functional rafts in cell membranes. *Nature* 387, 569–572
32 Kuehn, M.J. *et al.* (1998) COPII-cargo interactions direct protein sorting into ER-derived transport vesicles. *Nature* 391, 187–190
33 Grote, E. *et al.* (1995) A targeting signal in VAMP regulating transport to synaptic vesicles. *Cell* 81, 581–589
34 Salem, N. *et al.* (1998) A v-SNARE participates in synaptic vesicle formation mediated by the AP3 adaptor complex. *Nat. Neurosci.* 1, 551–556
35 Springer, S. and Schekman, R. (1998) Nucleation of COPII vesicular coat complex by endoplasmic reticulum to Golgi vesicle SNAREs. *Science* 281, 698–700
36 Springer, S. *et al.* (1999) A primer on vesicle budding. *Cell* 97, 145–148
37 Collins, J. and Mayblin, B. (1997) *Introducing Derrida*, Totem Books

Cell adhesion: old and new questions

Richard O. Hynes

Metazoans clearly need cell adhesion to hold themselves together, but adhesion does much more than that. Adhesion receptors make transmembrane connections, linking extracellular matrix and adjacent cells to the intracellular cytoskeleton, and they also serve as signal transducers. In this article, I briefly summarize our present understanding of the molecular basis and biological consequences of cell adhesion and discuss how our current knowledge sheds light on questions of specificity of cell adhesion. I offer some thoughts and speculations about the evolution of cell-adhesion molecules and processes, consider their inter-relationships with other forms of cell–cell communication and discuss unresolved questions ripe for investigation as we enter the postgenomic era.

Even a cursory consideration of metazoan anatomy and development forces the realization that the associations of cells in epithelia, their attachment to basement membranes and the migrations of cells and projections of neurons all require selective adhesion of cells to one another and to extracellular matrices (ECMs). Recognition of this requirement led to a spirited debate between proponents of a large number of highly selective adhesion receptors, and advocates of models in which quantitative differences in adhesive strength, without necessarily a large spectrum of individual specificities, were invoked to explain differential cell adhesion. Similarly, the phenomenon of induction, in which one tissue influences the developmental fate of adjacent tissues, clearly relies on cell–cell interactions, and experimental embryologists attempted to define whether induction relies on diffusible signals or on cell–cell or cell–matrix contacts. Neither the issue of specificity of cell adhesion nor the question of the mechanistic bases of induction could be resolved without molecular biology. Now, with the benefit of a couple of decades of molecular analysis, we can see that there is some truth to all of the earlier models. The specificity of cell adhesion comes from combinatorial expression and interactions among a large, but not unlimited, number of adhesion receptors, and induction relies on diffusible ligands binding to receptors, on cell–cell contacts and on cell–matrix adhesion. The distinctions among these three mechanisms are not actually that great – adhesion receptors signal much like receptors for growth factors and should be considered in parallel with them.

Before considering the biological functions of cell adhesion, we need to define the players. Figure 1 in Box 1 diagrams the structures of representative cell–cell adhesion receptors. Fortunately, many adhesion receptors fall into a relatively small number of families, the major ones being shown in Fig. 1. Other families of adhesion receptors, such as syndecans and other membrane-bound proteoglycans, the disintegrin family and others are less well understood at this time. In addition to their roles in binding cells to their neighbours (Fig. 1) or to ECM (Fig. 1), engagement of cell-adhesion receptors has major effects on many aspects of cell behaviour – cell shape and polarization, cytoskeletal organization, cell motility, proliferation, survival and differentiation. How do they accomplish all these functions?

Cytoskeletal connections

Crucial to the effects of adhesion receptors on intracellular organization and cell motility is the fact that their cytoplasmic domains connect to the cytoskeleton. Figure 1a shows how integrins bind to linker proteins, which in turn make direct and indirect connections to F-actin filaments, thus establishing a mechanical link between the fibrils of the ECM and the filaments of the cytoskeleton[9,10]. The connection of classic cadherins to the actin cytoskeleton that occurs at cell–cell junctions is analogous, although the molecules involved are different (Fig. 2a)[1,11,12]. Although integrins appear to be the major receptors for ECM, they are not the only ones. One well-studied example, of considerable interest because of its involvement in

Richard O. Hynes
rohynes@mit.edu

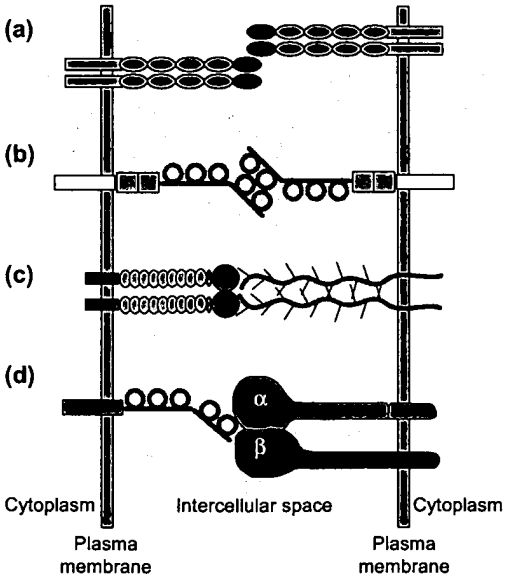
Howard Hughes Medical Institute and Center for Cancer Research, Dept of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

TCB, Vol. 9, N° 12 (0962-8924) PII: S0962-8924(99)01667-0
TIBS, Vol. 24, N° 12 (0968-0004) PII: S0968-0004(99)01479-6
TIG, Vol. 15, N° 12 (0168-9525) PII: S0168-9525(99)01891-0

BOX 1. Major classes of cell-adhesion receptors

(a) Cadherins

Cadherins are primarily and centrally involved in cell–cell adhesion (Fig. 1). The so-called classic cadherins (shown) currently number ~20 in vertebrates[1]. Their extracellular domains contain five characteristic cadherin repeats, each comprising a sandwich of β sheets. Cadherins mediate Ca^{2+}-dependent homophilic (like-with-like) adhesion between cells through the most distal cadherin repeats. Classic cadherins share homologous cytoplasmic domains that link to the actin cytoskeleton. Both structural and functional analyses suggest that the functional unit is a dimer as shown. As for other adhesion receptors, clustering of cadherins is important for their functions, and multiple dimer–dimer interactions are believed to provide sufficient local avidity to mediate cell–cell adhesion. Desmosomal cadherins (desmocollins, desmogleins), although related to classic cadherins in their extracellular domains, have distinct cytoplasmic domains that link to intermediate filaments. Other subclasses of the cadherin superfamily are known as protocadherins[2,3], and these typically have six cadherin repeats. Unlike classic and desmosomal cadherins, each of which is encoded by a separate genetic locus, protocadherins appear to be encoded by complex genetic loci with multiple (15–22) tandem exons each encoding one entire extracellular and transmembrane domain upstream of a single common cytoplasmic domain[3]. Each protocadherin subfamily is encoded by one such complex locus, but the mechanisms by which individual family members are generated remain unclear.



(b) Immunoglobulin superfamily

The second major class of adhesion receptors comprises the immunoglobulin superfamily (Ig-SF), characterized by the presence of varying numbers of Ig-related domains[4]. Like cadherin domains, these are sandwiches of two β sheets held together by hydrophobic interactions. This is a stable structure that occurs also in another domain common among adhesion molecules: fibronectin type III (Fn3) domains (boxes), which frequently occur in tandem with Ig domains (circles) in cell-adhesion receptors. Fn3 domains also occur in adhesive proteins of the extracellular matrix (ECM) such as fibronectin and tenascin and in the ligand-binding domains of cytokine receptors. Since homologous Ig/Fn3 receptors occur in insects, nematodes and vertebrates, this arrangement is clearly evolutionarily ancient. Indeed, these two domains probably originated in the context of cell-adhesion receptors early in metazoan evolution; their later appearance in immunoglobulins and fibronectin appears restricted to chordates.

The Ig superfamily is diverse, numbering well over 100 members in vertebrates. In addition to adhesion receptors containing both Ig and Fn3 repeats such as N-CAM (b), numerous molecules with one or more Ig domains play roles in cell–cell interactions in the immune system and elsewhere. Different Ig-SF members participate in homophilic interactions, as shown here for N-CAM, or in heterophilic interactions with other Ig-SF members, with integrins [see panel (d) and below] or with ECM proteins (e.g. DCC-netrins, see article by Tessier-Lavigne and Goodman in this issue). Where they have been mapped, the interaction sites typically are in the distal Ig domains. There are fewer data on dimerization, clustering and cytoskeletal connections than for cadherins, although some evidence suggests that such interactions also contribute to the functions of Ig-SF receptors.

(c) Selectins

Another well-studied group of cell adhesion receptors comprises the selectins and their counter-receptors[5,6]. The figure shows a heterophilic interaction between a selectin (P selectin) and its counterreceptor, a heavily glycosylated protein (PSGL-1). Binding is through the C-type lectin domain (pink) in the selectin, which recognizes specific carbohydrate groupings in the counter-receptor/ligand.

Unlike cadherins and Ig-SF members, which are evolutionarily ancient and widely expressed, selectins are currently known only in cells of the vertebrate circulation (endothelium and blood cells), although other lectins are widely distributed. Given the great potential for specificity that lies in carbohydrate structures, it seems likely that additional carbohydrate-specific receptors, such as galectins and the C-type lectins expressed by natural killer cells, will be increasingly recognized to be important.

Selectins and their ligands play a crucial role in the adhesion of leukocytes to endothelium, where their cooperation with integrins and Ig-SF receptors is one of the best-understood examples of cell-adhesion specificity, which arises from tightly regulated display and interaction among a limited number of receptors[5,6].

(d) Integrins

The final major family of adhesion receptors is the integrins[7,8]. Unlike all the others, these are heterodimers. In mammals, there are genes for eighteen α and eight β integrins; many α–β combinations fail to occur but at least two dozen are well defined. Most integrins are predominantly or exclusively receptors for ECM proteins such as fibronectins, laminins and collagens (Fig. 1a), but a few also play important roles in heterotypic cell adhesion, most notably of leukocytes, where they bind to counter-receptors of the Ig superfamily (ICAMs, VCAM-1, MAdCAM-1) or, in one case, a cadherin (αEβ7–E-cadherin). The figure shows a heterophilic interaction between an Ig-SF receptor (ICAM-1) and an integrin; the binding site is in the distal Ig repeats in ICAM-1 and partakes of both subunits in the integrin. Integrins play a central role in cell adhesion to basement membranes, in the polarization of cells induced by that adhesion and in cell migration upon and through ECM.

muscular dystrophies, is the dystroglycan complex, which connects dystrophin/actin inside the cell to laminin and/or agrin in the extracellular matrix (Fig. 1b)[13]. Although studied most extensively in muscle cells, analogous dystroglycan complexes clearly function in other cells.

Transmembrane structural connections, as shown in Fig. 1 and also demonstrated for other adhesion receptors [e.g. hyaluronan/CD44/ezrin–radixin–moesin (ERM) proteins], appear to be a common feature. There are preliminary indications that some immunoglobulin superfamily (Ig-SF) receptors also make cytoskeletal

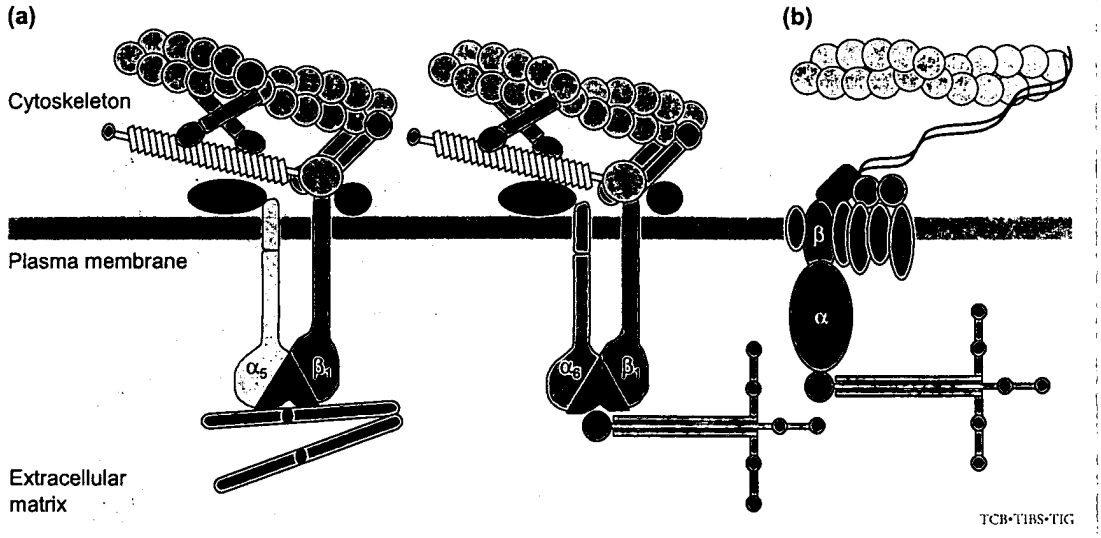


FIGURE 1. Transmembrane connections between the extracellular matrix (ECM) and the cytoskeleton. (a) Integrins (α5β1, α6β1) comprise the major receptors for ECM proteins, such as fibronectin (pink) and laminins (green), as shown here. Their extracellular domains bind to specific sites in the ECM proteins. Their cytoplasmic domains bind to submembranous cytoskeletal proteins such as talin (yellow) and α-actinin (lilac) and, through them, to other linkers such as vinculin (brown) or to actin microfilaments (grey). Additional cytoplasmic proteins (red) are also recruited; many of these function in signalling (see Fig. 2b). (b) Dystroglycan (αβ) together with sarcoglycans (blue) form another transmembrane link to laminin or to agrin (not shown) and bind via dystrophin (black) to actin filaments. Classic cadherins also link to the actin cytoskeleton via catenins (see Fig. 2a). For both integrins and cadherins, variant members of the families with divergent cytoplasmic domains (α6β4 integrin, desmogleins and desmocollins) connect instead to intermediate filaments via desmoplakins and other linker proteins.

connections (e.g. N-CAM/fodrin, ICAMs/ERM proteins) and that selectins or their counter-receptors might make similar connections that lead to their clustering on microvilli. The connections to the cytoskeleton affect not only intracellular organization but also cell adhesion itself. The adhesive functions of integrins and cadherins depend upon these cytoskeletal connections. Some of this dependence is presumably related to the clustering necessary to provide sufficient local avidity for stable cell adhesion. However, at least for integrins and possibly for other adhesion receptors, there can be more to it than that. Connection to the cytoskeleton can 'activate' integrins, changing their conformation and increasing their ability to bind to ligands. This ability to control the affinity and/or avidity of integrins is crucial to proper cell adhesion and is known as 'inside-out' signalling[7,14]. As we will see in the next section, integrins and cadherins are in fact two-way signalling receptors, and the same might be true for most adhesion receptors.

Signal transduction by adhesion receptors

A fundamental advance in the past decade has been the demonstration that cell-adhesion receptors transduce signals. This is best understood for integrins, which display a repertoire of signal-transduction capabilities at least as diverse as most growth-factor receptors (Fig. 2b)[14–16]. Their effects include activation of Rho-family GTPases leading to changes in cytoskeletal organization, activation of mitogen-activated protein (MAP) kinase pathways and activation of an array of protein and lipid kinases. These signalling pathways allow integrins to influence cell-cycle progression, cell survival and gene expression in addition to their effects on cell adhesion and morphology. In fact, most cells will not proliferate or survive unless they are adhering to a substrate – so-called anchorage dependence. Provision of soluble growth factors such as epidermal growth factor (EGF) or platelet-derived growth factor (PDGF) is not sufficient; input from integrin signalling is also necessary, and there is considerable crosstalk and cooperation between integrins and growth-factor receptors. This cooperation occurs at many levels, ranging from membrane-proximal interactions, in which the different types of receptor influence each other's activity, to multiple inputs into common pathways. Indeed, it is not realistic to consider either adhesion receptors or growth-factor receptors separately – they are part of an integrated system.

This integration is clearly demonstrated by the cadherin/β-catenin system[11,12]. β-catenin is a cytoskeletal connector of classic cadherins, but it is also a central player in signal transduction, functioning as a transcriptional activator whose levels are elevated in response to Wnt signalling (Fig. 2a). The interplay between cell–cell adhesion and the Wnt signalling pathway is complex, with each affecting the other, just like the interplay between integrins and tyrosine kinase receptors. Other members of the cadherin superfamily presumably affect different signalling pathways; protocadherins fall into subfamilies, each with a distinct cytoplasmic domain, and one protocadherin subclass was first identified by its interactions with the Src-family kinase Fyn[2].

It is also becoming clear that integrins, at least, do not signal by themselves; they are frequently associated with accessory transmembrane molecules (tetraspanins, CD47, caveolin, syndecans) that contribute to the diversity of their signalling capacities[17]. It is possible to draw an analogy with the well-analysed T- and B-cell receptors and their multiple associated signalling molecules[18,19]. There are also indications that other adhesion receptors function as constituents of complexes involving multiple signalling molecules. One can readily extrapolate from the current data and postulate that most or all signal transduction relies on associations among multiple receptors, including both adhesion receptors and receptors for soluble ligands.

A receptor continuum: soluble ligands to ECM to cell–cell contact

There is, in fact, little or no justification for drawing a distinction between adhesion receptors and receptors for soluble ligands; both signal, often affecting the same signal-transduction pathways. Indeed many 'soluble' growth factors often do not function as truly soluble

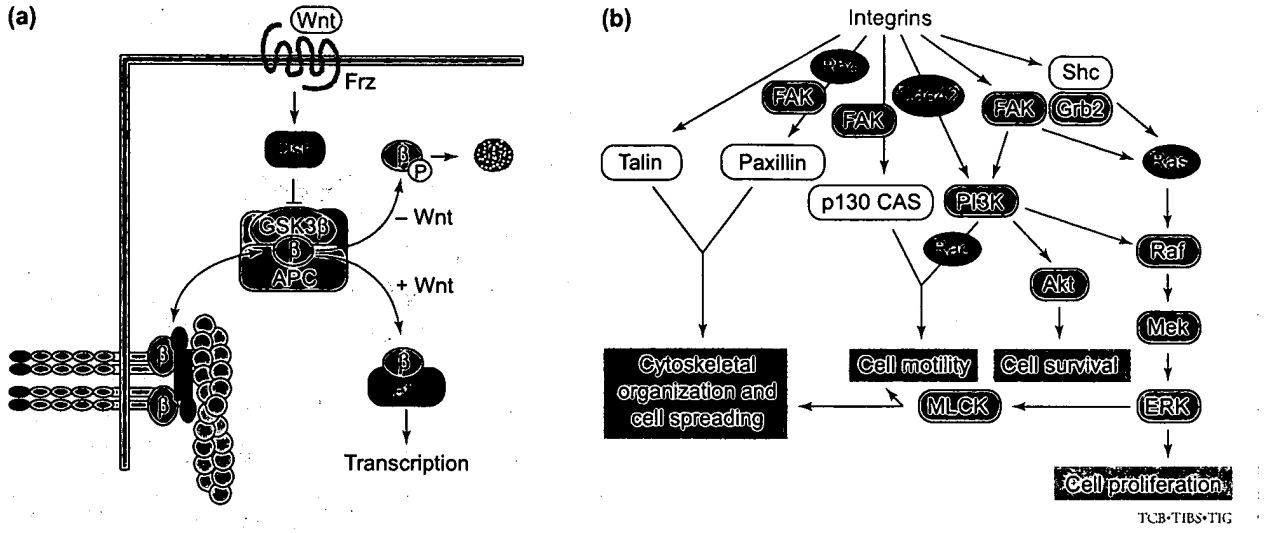


FIGURE 2. Signalling mediated by adhesion receptors. (a) Classic cadherins bind to β-catenin through their cytoplasmic domains. β-catenin can link via α-actinin to the actin cytoskeleton or it can bind to a large protein complex containing adenomatous polyposis coli (APC) and the serine/threonine kinase glycogen synthase kinase 3β (GSK3β). The latter phosphorylates β-catenin, targeting it for degradation by the proteasome. Wnt binding to its receptor, Frizzled (Frz), leads to inhibition of GSK3β, allowing β-catenin to accumulate and bind to the transcription factor Lef-1/TCF. The β-catenin–Lef-1 complex moves to the nucleus and activates transcription. Thus, the balance between cadherin association, degradation and Wnt signalling controls the level of β-catenin–Lef-1. (b) Integrins activate a large array of signalling intermediates, including small GTPases (red), protein kinases (green), cytoskeletal proteins (yellow) and others. Acting through these intermediates, which can also be activated by various growth-factor receptors, integrins can greatly affect many biological responses (blue boxes). Abbreviations: FAK, focal-adhesion kinase; MLCK, myosin light chain kinase; PI3K, phosphoinositide 3-kinase.

molecules. Many (transforming growth factor β, fibroblast growth factors, Wnts, Hedgehogs) bind in one way or another to the ECM and are presented to their signal-transduction receptors as insoluble mediators. The whole concept of morphogenetic gradients incorporates the idea that morphogens are both soluble and anchored. So the boundary between soluble ligands and ECM ligands is blurred. Similarly, receptors that mediate cell–cell contacts such as the T-cell receptor[18] have much in common with those binding soluble or bound antigen or antibody (B-cell receptor, Fc receptor)[19]. Receptor pairs such as the eph/ephrin[20] and Notch/Delta/Serrate families[21], Sevenless/Boss and receptor tyrosine phosphatases[22] all share domains and signal-transduction mechanisms, or both, with growth factors, ECM or classical growth-factor receptors. In some cases, these receptors have been shown to mediate cell–cell adhesion. In other words, there is considerable commonality of evolution and function among the different types of receptors.

If we return to the question of embryonic induction first raised 70 years ago by experimental embryologists and reconsider the debate as to whether induction relies on soluble factors, extracellular matrix or cell–cell contact, that question now seems somewhat moot. The answer is that all three can, and typically do, contribute, but they are part of a continuum, and all feed into a common network of intracellular signals with much synergy and crosstalk among them. A major challenge ahead of us is to understand the integration of all these inputs to generate coherent responses.

Where do we stand and where do we go from here?

Given what we now know about adhesion receptors, what can we say about the specificity of cell adhesion? Is it due to a very large number of receptors, sufficient for example to confer identity on each retinal axon or synapse? How many adhesion receptors are there in the genome? With the sequence of the first metazoan genome, that of *Caenorhabditis elegans* [see articles in *Science* (1998) 282, 2011–2046], we can begin to answer these questions – some of the answers are surprising.

One striking result from the *C. elegans* sequence is the discovery of a very large number of genes that encode ECM proteins. What are all these proteins for? They could serve purely structural roles or act as docking sites for presentation of growth factors, gradients of morphogens or chemoattractants. The ECM performs such functions in vertebrates, and even well-studied matrix proteins such as fibronectin, tenascin and agrin contain many highly conserved segments whose functions remain completely obscure. There is clearly a great deal that we do not understand about the functions of the ECM. Will a similar plethora of putative matrix proteins emerge from the fly and vertebrate genome sequences? There is every reason to believe that they will; the discovery of new matrix proteins continues apace even before the flood of genomic sequence data. One recent example is the discovery of netrins[23] as axonal guidance molecules. The large number of matrix proteins is not matched by a large number of integrins. Does this mean that the integrins are very promiscuous, that other matrix receptors exist or that these putative matrix proteins do not interact directly with cells?

There appear to be only two integrins in *C. elegans*. Strikingly, the two integrin α genes appear related to two distinct subfamilies of vertebrate integrins, one that binds to laminins (α3, α6, α7) and one that binds to proteins containing the sequence RGD, such as fibronectin and vitronectin (α5, α8, αv, αIIb)[24]. *Drosophila* also contains clear representatives of each of these two integrin subfamilies[24]. Thus, these two subfamilies apparently evolved prior to the divergence of nematodes, arthropods and chordates. The same is true for laminin and type IV collagen, although not for fibronectin, which is absent from nematodes (and apparently also flies) and might be a vertebrate invention. It is plausible to argue that some very early metazoan evolved laminin and collagen to build a basement membrane and integrins for cells to attach to this membrane. During evolution, vertebrates have acquired multiple integrin genes. How many more will we find when the *Drosophila*, human and mouse

genomes are sequenced in the next few years? The limited repertoire in *C. elegans* might suggest not many. On the other hand, the number of known cadherin/ protocadherin genes has more than doubled in just the past year with the application of human genomic analyses. This provides a glimpse of what might be just around the corner.

The *C. elegans* genome has 18 genes that contain cadherin repeats; we already know of more than 70 in humans, and the number is rising fast. Why do we need so many more integrins and cadherins than worms do? One obvious suggestion might be the elaboration of our nervous system; many cadherins and protocadherins are expressed in the brain, apparently differentially in different brain regions or in individual neurons[2,25]. Could they provide selectivity in neuronal or synaptic adhesion along the lines of the chemoaffinity hypothesis proposed 60 years ago by Sperry[26]? Both classic and protocadherins, as well as integrins, are expressed at synapses[27–29]. The recent discovery of multiple genes encoding protocadherins raises the exciting possibility that a large number of adhesion receptors confer synaptic selectivity. If these protocadherins can form heterodimers or heteromultimers, then the number of potential combinations becomes very large[2]. The tantalizing organization of the protocadherin loci, with multiple variable exons and a common constant region is reminiscent of immunoglobulins or T-cell receptors[3]. There is currently no evidence for DNA rearrangements at these loci, although mutations in some genes responsible for repair of double-strand breaks lead to selective apoptosis of early postmitotic neurons, encouraging speculation[30]. Even if DNA rearrangements were to occur, there is as yet no sign of the multiple combinatorial variation seen in the immune system. Nonetheless, the existence of >50 genes for protocadherins (conceivably 2500 heterodimers) offers a fair degree of variation.

Our current picture of leukocyte adhesion to the endothelium offers a good example of how a high degree of specificity in cell adhesion can be generated using only a limited number of not particularly selective adhesion receptors[5,6]. Three selectins and their ligands, three to five integrins and five to six Ig-SF receptors appear to be sufficient to target leukocytes specifically to multiple sites during inflammation or lymphocyte trafficking. This selectivity relies on tightly regulated expression and, importantly, on activation of the integrins through crosstalk from selectins and chemokine receptors. The specificity therefore relies more on spatiotemporal regulation, combinatorial expression and activation of several receptors than on the intrinsic specificity of individual receptors.

Therefore, in considering how to explain the specificity of cell–cell adhesion, we have a fairly large number of receptors (hundreds), and we will soon know exactly how many. Combinatorial display and the ability of these receptors to cooperate with each other and with 'classical' signalling receptors and to be fine-tuned in terms of their state of activation could provide enough potential spatiotemporal specificity. The challenge now will be to exploit our knowledge of the list of players to understand the complexity of individual biological systems.

While questions arising from developmental biology represented one impetus to understand cell adhesion, others came from a desire to understand pathological processes. Altered adhesion properties were recognized early as a feature of cancer cells, and the tightly regulated adhesion of blood cells is central to haemostasis, thrombosis, leukocyte trafficking and inflammation. A satisfying outcome of cell-adhesion research has been the discovery that most cell-adhesion events, be they developmental, physiological or pathological, rely on members of a limited number of families of cell-adhesion receptors. This realization has led to a very productive synergy among the originally separate areas of investigation. Molecular analyses of cell adhesion have revealed that adhesion has profound effects on cells that go far beyond merely sticking them together. Furthermore, detailed understanding of cell-adhesion receptors has opened the way to manipulating their functions, leading to therapeutic strategies applicable to pathological processes involving cell adhesion.

References

1 Yap, A.S. *et al.* (1997) Molecular and functional analysis of cadherin-based adherens junctions. *Annu. Rev. Cell Dev. Biol.* 13, 119–146
2 Kohmura, N. *et al.* (1998) Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* 20, 1137–1151
3 Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779–790
4 Walsh, F.S. and Doherty, P. (1997) Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. *Annu. Rev. Cell Dev. Biol.* 13, 425–456
5 Lasky, L.A. (1995) Selectin–carbohydrate interactions and the initiation of the inflammatory response. *Annu. Rev. Biochem.* 64, 113–139
6 Kansas, G.S. (1996) Selectins and their ligands: current concepts and controversies. *Blood* 88, 3259–3287
7 Hynes, R.O. (1992) Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* 69, 11–25
8 Hemler, M.E. (1999) Integrins. In *Guidebook to the Extracellular Matrix and Adhesion Proteins* (Kreis, T. and Vale, R., eds), Sambrook and Tooze Publishers, Oxford University Press
9 Burridge, K. and Chrzanowska-Wodnicka, M. (1996) Focal adhesions, contractility, and signaling. *Annu. Rev. Cell Dev. Biol.* 12, 463–518
10 Jockusch, B.M. *et al.* (1995) The molecular architecture of focal adhesions. *Annu. Rev. Cell Dev. Biol.* 11, 379–416
11 Barth, A.I., Nathke, I.S. and Nelson, W.J. (1997) Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways. *Curr. Opin. Cell Biol.* 9, 683–690
12 Ben-Ze'ev, A. and Geiger, B. (1998) Differential molecular interactions of beta-catenin and plakoglobin in adhesion, signaling and cancer. *Curr. Opin. Cell Biol.* 10, 629–639
13 Henry, M.D. and Campbell, K.P. (1996) Dystroglycan: an extracellular matrix receptor linked to the cytoskeleton. *Curr. Opin. Cell Biol.* 8, 625–631
14 Schwartz, M.A. *et al.* (1995) Integrins: emerging paradigms of signal transduction. *Annu. Rev. Cell Dev. Biol.* 11, 549–599
15 Clark, E.A. and Brugge, J.S. (1995) Integrins and signal transduction pathways: the road taken. *Science* 268, 233–239
16 Giancotti, F.G. (1997) Integrin signaling: specificity and control of cell survival and cell cycle progression. *Curr. Opin. Cell Biol.* 9, 691–700
17 Hemler, M.E. (1998) Integrin associated proteins. *Curr. Opin. Cell Biol.* 10, 578–585
18 Qian, D. and Weiss, A. (1997) T cell antigen receptor signal transduction. *Curr. Opin. Cell Biol.* 9, 205–212
19 Isakov, N. (1997) Immunoreceptor tyrosine-based activation motif (ITAM), a unique module linking antigen and Fc receptors to their signaling cascades. *J. Leukocyte Biol.* 61, 6–16
20 Pasquale, E.B. (1997) The Eph family of receptors. *Curr. Opin. Cell Biol.* 9, 608–615
21 Kimble, J. and Simpson, P. (1997) The LIN-12/Notch signaling pathway and its regulation. *Annu. Rev. Cell Dev. Biol.* 13, 333–361
22 Neel, B.G. and Tonks, N.K. (1997) Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* 9, 193–204
23 Culotti, J.G. and Merz, D.C. (1998) DCC and netrins. *Curr. Opin. Cell Biol.* 10, 609–613
24 Stark, K.A. *et al.* (1997) A novel alpha integrin subunit associates with betaPS and functions in tissue morphogenesis and movement during *Drosophila* development. *Development* 124, 4583–4594
25 Arndt, K. *et al.* (1998) Cadherin-defined segments and parasagittal cell ribbons in the developing chicken cerebellum. *Mol. Cell. Neurosci.* 10, 211–228
26 Sperry, R.W. (1963) Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. U. S. A.* 50, 703–710
27 Uchida, N. *et al.* (1996) The catenin/cadherin adhesion system is localized in synaptic junctions bordering transmitter release zones. *J. Cell Biol.* 135, 767–779
28 Fannon, A.M. and Colman, D.R. (1996) A model for central synaptic junctional complex formation based on the differential adhesive specificities of the cadherins. *Neuron* 17, 423–434
29 Nishimura, S.L. *et al.* (1998) Synaptic and glial localization of the integrin alphavbeta8 in mouse and rat brain. *Brain Res.* 791, 271–282
30 Gao, Y. *et al.* (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* 95, 891–902

Acknowledgements

I thank Denisa Wagner for helpful criticism of the text, Colleen Leslie for editorial assistance and the NIH and the Howard Hughes Medical Institute for support.

Minireview

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 41, Issue of October 14, pp. 25235–25238, 1994

Integrin-mediated Cell Adhesion: The Extracellular Face*

Joseph C. Loftus, Jeffrey W. Smith, and Mark H. Ginsberg‡

From the Department of Vascular Biology, The Scripps Research Institute, La Jolla, California 92037

Cell adhesion regulates embryonic development by controlling cell migration, growth, and differentiation. Additionally, adhesion contributes to the processes of malignant transformation, inflammation, hemostasis, and immune recognition (1, 2). Cell adhesive events are mediated by transmembrane receptors that belong to a limited number of supergene families. These include the integrins (3, 4), immunoglobulin supergene family (5), cadherins (6), selectins (7, 8), CD44-related molecules (9, 10), and transmembrane proteoglycans (11). The role of each of these supergene families as general cell adhesion molecules has only been appreciated in the last 10 years, yet cell adhesion has now become an area of intensive investigation. This rapid progress precludes a comprehensive review of this entire field. We will focus on the integrin family, since the key insights concerning these receptors are likely to have some relevance to other adhesion receptors.

All integrins discovered to date are heterodimers of α and β subunits, which are products of separate genes (2, 4, 12, 13) and are mutually interdependent for correct processing and surface expression (14). To date, the integrin family is composed of 14 α subunits and 8 β subunits. Integrins possess a generally conserved structure: a large extracellular domain formed by both the ~1000-residue α and ~750-residue β subunits and a transmembrane segment from each subunit. In general, each subunit possesses a short cytoplasmic C-terminal tail (15–17) with the striking exception of the β_4 subunit whose cytoplasmic domain is more than 1000 residues (18–20). When the integrin gene family was first proposed in 1987 (21, 22), a simple vertical organization consisting of three β subunits that formed unique heterodimers with distinct α subunits was envisaged. The α subunits, at the time, were thought to pair only with specific β subunits. With the discovery that the α subunits appear to have evolved independently of β subunits (23) and the discovery of novel β subunits (24–26), which form heterodimers with existing α subunits, and with the discovery that certain α subunits can form heterodimers with multiple β subunits (24, 26, 27) more complicated schemes have been required to characterize this family (4). To date, all integrin gene disruptions are associated with evident phenotypes. These findings unambiguously establish the critical biological role of integrins (28–33). In general, integrins perform their functions by interacting with components of the insoluble extracellular matrix or surface proteins on other cells to form links with intracellular elements involved in bidirectional signaling events. These signaling events and the integrin cytoplasmic domains that mediate many of them have been extensively reviewed recently (4, 12, 15–17, 34–37). Thus, we will focus on the recognition of adhesive ligands by the extracellular domain of integrins.

Integrins bind to diverse ligands including components of the extracellular matrix (38–40), cell surface Ig superfamily receptors (41), components of microorganisms (42–44), and certain plasma proteins (45, 46). In most cases, these interactions are divalent cation-dependent, and mapping of integrin recognition sequences in ligands almost invariably identifies an acidic residue as a key component (47–55). In some protein ligands, these recognition sites can be assigned to short linear peptide sequences, *e.g.* Arg-Gly-Asp (48). Even in these cases, additional discontinuous regions of the protein may be involved in ligand recognition as demonstrated in fibronectin (56–58) (Fig. 1). Indeed, the general theme seems to be that integrins will recognize short peptide sequences often presented in extended loops containing β turns (59–63). Other regions of protein ligands may then contribute secondary interactive sites.

There are also emerging generalities about the sites in integrins that recognize ligands. Electron microscopy of rotary shadowed preparations of purified integrins suggests that these receptors are comprised of a globular ~10-nm head (64, 65) to which ligands bind (66) and two rodlike tails probably containing C-terminal portions of α (66) and β (67) subunits and their transmembrane domains (68). Biophysical analyses of integrins in detergent solution agree, in general, with the idea that these are asymmetric molecules (69, 70). In addition, Calvete and co-workers (71–76) have examined the disulfide bond arrangements and intersubunit contacts in proteolytic fragments of a prototype integrin, $\alpha_{IIb}\beta_3$. Rocco and co-workers (77) made an ingenious effort to accommodate these findings to the biophysical and electron microscopic data. Integrins are conformationally labile (78–81) and are subject to proteolysis (82) and disulfide bond rearrangement (83). Thus the purification method and storage of the receptor (84, 85) should be considered in evaluating studies from different groups.

Potential ligand binding sites in integrins have been investigated through a combination of immunological, biochemical, and mutational approaches. Proteolysis (86) or expression of recombinant truncated (87) $\alpha_{IIb}\beta_3$ produced ligand binding fragments that contain at least the N-terminal half of the α and β subunits. These results are in good agreement with previous cross-linking studies that suggested that ligand recognition sites reside in the N-terminal portion of both α_{IIb} (88, 89) and β_3 (88, 90, 91) subunits. These results also support the concept that high affinity ligand recognition requires both subunits (84, 92–96) and consequently may involve multiple ligand contact points. Several such potential contact points have now been identified.

Chemical cross-linking of an RGD peptide to $\alpha_{IIb}\beta_3$ followed by proteolytic digestion and amino acid sequencing indicated that a 72-residue sequence in β_3 (Asp^{109}–Glu^{171}) is proximal to bound ligand (91). This localization was supported by identification of the overlapping region of β_3 (Glu^{65}–Glu^{220}) by photoaffinity cross-linking of an RGD peptide to $\alpha_v\beta_3$ (90). This region of β_3 probably is directly involved in ligand recognition because 1) point mutations in the region abrogate ligand binding function (97, 98), 2) certain antibodies directed against this region inhibit ligand binding (99–101), and 3) a gain of ligand binding function mutation involves this region (102). This region of β_3 is highly conserved among integrin classes, suggesting that this is a common ligand contact site in all integrins. That hypothesis has been supported by the loss of ligand binding function associated with mutations at residues homologous to Asp^{119} in β_1 (103) and β_6 (104).

The prototype mutation in this ligand contact site, β_3 (D119Y), also alters the conformation of $\alpha_{IIb}\beta_3$ in a manner consistent with loss of bound divalent cation (105). Moreover, Asp^{119} resides within a region of β_3 enriched in amino acids with oxygenated side chains (Asp^{119}, Ser^{121}, Ser^{123}, Asp^{126}, Asp^{127}, and Ser^{130}) whose linear spacing approximates that of the oxygenated residues in the calcium binding loop of EF-hand proteins (106) suggesting that these residues may provide coordinating ligands for divalent cations. A synthetic β_3 peptide corresponding to residues 118–131 directly binds terbium, a luminescent calcium analog. Moreover, substitution of Asp^{119} by alanine in this peptide reduced the peptide/terbium (107) interaction, supporting involvement of this region in both ligand recognition and cation binding. Alanine substitutions in β_3 Asp^{119}, Ser^{121}, or Ser^{123} resulted in deficits in the binding of both macro-

* This minireview will be reprinted in the Minireview Compendium, which will be available in December, 1994. This work was supported by National Institutes of Health Grants HL48728 (to J. C. L. and M. H. G.), AR27214 (to M. H. G.), HL28235 (to M. H. G.), HL42977 (to J. C. L.), and CA56483, AR42750, and TRDRP 3RT-0258 (to J. W. S.). This work was done during the tenure of an established investigatorship of the American Heart Association (J. C. L. and J. W. S.). This is publication 8702-VB from the Scripps Research Institute.

‡ To whom correspondence should be addressed. Tel.: 619-554-7124; Fax: 619-554-6403; E-mail: ginsberg@scripps.edu.

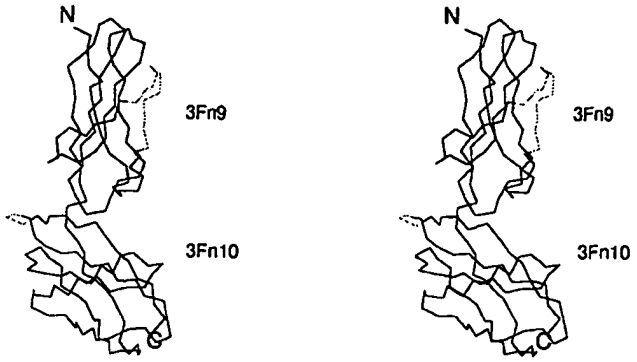


FIG. 1. **α carbon backbone representation of a model of the fibronectin 9th and 10th type III repeats.** The central cell binding domain of fibronectin is comprised of ~90-residue repeats designated "type III repeats." Integrin $\alpha_{IIb}\beta_3$ recognizes the 9th (*3Fn9*) and 10th (*3Fn10*) type III repeats. Depicted is a stereo view of the backbone of a fibronectin 3Fn(9–10) structural model generated as described (58). Distinct, linear peptide sequences that regulate fibronectin-integrin interaction have been identified within each of these fibronectin type III repeats. The Asp^{1373}–Thr^{1383} sequence in the 3Fn9 repeat and the Arg^{1493}–Asp^{1496} sequence in the 3Fn10 repeat are shown as *dotted lines*. The N terminus (*N*) and the C terminus (*C*) are shown. Note that each recognition sequence resides within an extended loop. (Reprinted, with permission, from Ref. 58.)

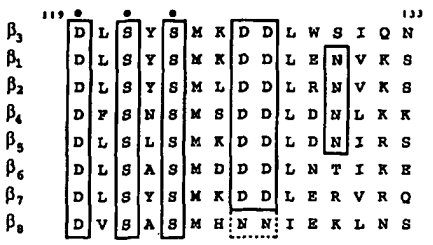


FIG. 2. **Alignment of the region of the putative β_3 ligand binding site with the deduced sequences of the other integrin β subunits.** Alanine substitution of indicated residues in β_3 (●) results in loss of ligand binding function. *Boxed* are highly conserved amino acids with oxygenated side chains with a linear spacing similar to the oxygenated residues in the calcium binding loop of EF-hand proteins (106). (Reprinted, with permission, from Bajt and Loftus (98).)

molecular and peptide ligands. In contrast, substitutions at positions Asp^{126}, Asp^{127}, or Ser^{130} did not affect ligand binding (98). These results further implicate this region of β_3 in ligand binding function of $\alpha_{IIb}\beta_3$ and assign functional roles to Asp^{119}, Ser^{121}, and Ser^{123}. This cluster of three oxygenated residues is also highly conserved among the integrin β subunits (Fig. 2). The divalent cation dependence of integrin function and the high degree of conservation of the functionally important oxygenated residues suggest that ligands interact with divalent cations bound to this highly conserved site in the β subunit (97). This hypothesis is further supported by the presence of critical oxygenated residues in integrin ligands (47–55) and the evidence that ligand binding to integrins may displace divalent cations (80, 108).

A second highly conserved potential ligand interactive site in β_3 was identified through a synthetic peptide corresponding to β_3 (211–222) that bound to fibrinogen and blocked its binding to $\alpha_{IIb}\beta_3$ (109). Antibodies directed against this peptide also inhibited the binding of adhesive proteins to the purified receptor (109). The mechanism of action of the synthetic β_3 (211–222) has been questioned, since the peptide also binds specifically to $\alpha_{IIb}\beta_3$ (110). Furthermore, two natural mutations at β_3 Arg^{214} (111, 112) result in loss of ligand binding function. Interpretation of that result is clouded by the finding that these mutations can also impair the stability of the $\alpha_{IIb}\beta_3$ heterodimer (113).

α subunit ligand contact sites have also been identified by chemical cross-linking approaches. Ligand-mimetic peptides cross-linked to the N-terminal region of α_{IIb} (114) and α_v (89) subunits. This localization was refined to a 21-amino acid stretch of α_{IIb} defined by Ala^{294}-Met^{314} (114), a region spanning the second putative divalent cation binding repeat of α_{IIb}. Peptides from this region and antibodies against them are reported to inhibit fibrinogen binding to $\alpha_{IIb}\beta_3$, supporting a role in ligand recognition (115). In addition, a synthetic peptide corresponding to α_{IIb} (300–312) inhibits clot retraction and platelet aggregation and directly binds fibrinogen (116). The immediate proximity of this peptide to the α_{IIb} (296–306) sequence further substantiates the importance of this region in the ligand binding function of the receptor. Finally, a recombinant fragment of α_{IIb}, spanning all four putative cation binding sites, has been reported to bind to cations and to fibrinogen (117). Mutational analyses of the cation binding sites in the α subunits have been hampered by the fact that some of these mutations block receptor expression (116, 118). Nevertheless, mutational evidence for a role of the cation binding site in the α_4 subunit has appeared (116). The involvement of the cation binding sites in α_4 in ligand binding suggests that these regions, like β_3 (109–171), may function in a general mechanism of ligand binding to integrins. Recent studies with α_L lend credence to this idea (119).

At least six integrin α subunits contain an additional ~200-residue sequence in their N-terminal third (120, 121). This sequence appears to have been inserted by exon shuffling (122–125); hence it is sometimes referred to as the I (Inserted) domain, and it is homologous to the A domains of von Willebrand factor (126–130). There is now growing evidence for the functional importance of this domain in ligand binding. Function-altering antibodies map to I domains of $\alpha_M\beta_2$ (Mac-1, CD11b/CD18) (131, 132) and $\alpha_2\beta_1$ (VLA-2) (133). Moreover, mutations in these domains block ligand binding function (132, 133), and an isolated I domain binds ligands (132). The I domain of α_M also binds cations (132). Moreover, there are critical oxygenated residues in the I domain for both cation and ligand binding (132, 133). Alignment of a region containing a functional Asp in the I domains of α_M and α_2 (132, 133) with the proposed ligand and cation binding site in the β_3 subunit reveals striking similarities (Fig. 3). From this alignment, a conserved motif of $D(\Phi)_5DXSXS\Phi$, where Φ is any hydrophobic residue and *X* is any residue, seems evident. In view of the sequence divergence in flanking residues, this similarity probably arose by convergence, possibly driven by a common function of divalent cation-dependent interactions. Further mutational analysis and evaluation of this motif in some of the other proteins in which it is found (*e.g.* cartilage matrix protein (134), type VI collagen (135), and thrombospondin-related anonymous protein (136)) should test this hypothesis.

As noted, the evidence for the role of the I domain in ligand binding is compelling. It is notable that I domain integrins manifest some differences in divalent cation preferences (137–140). Moreover, although some inhibitory peptides have been identified for these integrins (141–143) they appear to be of relatively low affinity. This should be contrasted with the high affinity small ligands that have been found for some other integrins (144–150). Thus it is possible that the I domain integrins have a distinct binding mechanism from other integrins. This possibility should be readily tested by mutational analysis of the highly conserved ligand binding site in β subunit partners of I domain-containing α subunits.

All of the ligand-integrin interactions discussed above involve divalent cations. Divalent cations could induce conformational changes exposing the integrin's ligand binding surfaces in a manner analogous to troponin C (151). Alternatively, the divalent cation may be part of the active site. As noted above, the structures of the proposed cation binding motifs in integrins and the sequences of integrin ligands suggest that they could simultaneously coordinate a single divalent cation. The cation binding motifs in the α subunits lack the 12th residue in the EF-hand calcium binding loops (152). Thus, Asp or Glu residues within the ligand could provide the final acidic residue to the cation coordination loop of the integrins (153, 154). Recently, quantitative analysis revealed that each $\alpha_{IIb}\beta_3$ possesses three Mn^{2+} binding sites; however, addition of synthetic peptide ligand resulted in displacement of manganous ions from two of these sites (108). This result is consistent with the finding that ligand binding to integrins results in conformational changes similar to those seen following chelation of divalent cations (80, 105). It is possible that during ligand binding a transient ternary complex is formed, followed by additional receptor-ligand contacts and displacement of the cation from the integrin. Additional support for this idea comes from the observation that when Co(II) bound to β_3 integrins is oxidized to Co(III), an inert form of the ion that cannot

β_3	D	L	Y	Y	L	M	D	L	S	Y	S	M	K	D	D	L
β_1	D	L	Y	Y	L	M	D	L	S	Y	S	M	K	D	D	L
β_2	D	I	Y	Y	L	M	D	L	S	Y	S	M	L	D	D	L
β_4	D	L	Y	I	L	M	D	F	S	N	S	M	S	D	D	L
β_5	D	L	Y	Y	L	M	D	L	S	L	S	M	K	D	D	L
β_6	D	L	Y	Y	L	M	D	L	S	A	S	M	D	D	D	L
β_7	D	L	Y	Y	L	M	D	L	S	S	S	M	K	D	D	L
β_8	D	L	Y	Y	L	V	D	V	S	A	S	M	H	N	N	I
α_1	D	I	V	I	V	L	D	G	S	N	S	I				
α_2	D	V	V	V	V	C	D	E	S	N	S	I				
α_L	D	L	V	F	L	F	D	G	S	M	S	L				
α_X	D	I	A	F	L	I	D	G	S	G	S	I				
α_X	D	I	V	F	L	I	D	G	S	G	S	I				
CON	D	Φ	Φ	Φ	Φ	Φ	D	x	S	x	S	Φ				

FIG. 3. **Conservation of potential integrin ligand binding domains.** Alignment of relevant integrin I domain and β subunit sequences demonstrates remarkable conservation and suggests a potential new motif involved in cation-dependent protein-protein interaction. Residues where mutations block ligand binding function are *boxed*. *CON*, consensus; Φ, any hydrophobic residue; *X*, any residue.

exchange or be displaced, ligand binding is inactivated (155). One important prediction is that a divalent cation should be able suppress ligand binding. Such cation-mediated suppression has been reported for several integrin-ligand pairs (137–140).

Integrin extracellular domains have masses in excess of 200 kDa. Consequently, it is not surprising that they may also undergo cation-independent binding interactions. Plasminogen binding to $\alpha_{IIb}\beta_3$ was the first example of cation-independent integrin ligand (157). Moreover, plasminogen also bound to $\alpha_{IIb}\beta_3$ (β_3 D119Y), a mutant that fails to bind to other ligands (157). Since plasminogen activators colocalize with integrins in focal adhesions (158), this interaction may facilitate local plasmin formation at these membrane microdomains. Further, many of the two-chain α subunit integrins are posttranslationally cleaved probably resulting in a C-terminal lysine on their heavy chain (159, 160), presenting a likely site for plasminogen binding. In addition, there is now abundant evidence that certain integrins concentrate in cell-cell junctions and may mediate cell-cell interactions (161–163), in some cases through homophilic binding (164, 165). Further, there are integrin-dependent cell-cell interactions that do not appear to be mediated through "classical" integrin-dependent recognition mechanisms (156, 166–169). Thus, there will probably be new surprises and new insights arising from the continuing analysis of integrin-ligand interactions.

Acknowledgments—We acknowledge our colleagues C. Buck, M. Hemler, R. Hynes, and T. A. Springer for their insightful suggestions and critical review.

REFERENCES

1. Edelman, G. M., and Crossin, K. L. (1991) *Annu. Rev. Biochem.* **60,** 155–190
2. Albelda, S. M., and Buck, C. A. (1990) *FASEB J.* **4,** 2868–2880
3. Ruoslahti, E. (1991) *J. Clin. Invest.* **87,** 1–5
4. Hynes, R. O. (1992) *Cell* **69,** 11–25
5. Buck, C. A. (1992) *Semin. Cell Biol.* **3,** 179–188
6. Takeichi, M. (1991) *Science* **251,** 1451–1455
7. Lasky, L. A. (1992) *Science* **258,** 964–969
8. McEver, R. P. (1992) *Curr. Opin. Cell Biol.* **4,** 840–849
9. Kincade, P. W., He, Q., Ishihara, K., Miyake, K., Lesley, J., and Hyman, R. (1993) *Curr. Top. Microbiol. Immunol.* **184,** 215–222
10. Herrlich, P., Zoller, M., Pals, S. T., and Ponta, H. (1993) *Immunol. Today* **14,** 395–399
11. Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T., Spring, J., Gallo, R. L., and Lose, E. J. (1992) *Annu. Rev. Cell Biol.* **8,** 365–393
12. Ginsberg, M. H., O'Toole, T. E., Loftus, J. C., and Plow, E. F. (1992) *Cold Spring Harbor Symp. Quant. Biol.* **57,** 221–231
13. Akiyama, S. K., Nagata, K., and Yamada, K. M. (1990) *Biochim. Biophys. Acta* **1031,** 91–110
14. O'Toole, T. E., Loftus, J. C., Plow, E. F., Glass, A., Harper, J. R., and Ginsberg, M. H. (1989) *Blood* **74,** 14–18
15. Sastry, S. K., and Horwitz, A. F. (1993) *Curr. Opin. Cell Biol.* **5,** 819–831
16. Hemler, M. E., Weitzman, J. B., Pasqualini, R., Kawaguchi, S., Kassner, P. D., and Berdichevsky, F. B. (1994) in *The Integrin* (Takada, Y., ed) CRC Press, Inc., Boca Raton, FL
17. Williams, M. J., Hughes, P. E., O'Toole, T. E., and Ginsberg, M. H. (1994) *Trends Cell Biol.* **4,** 109–112
18. Hogervorst, F., Kuikman, I., Von dem Borne, A. E. G., and Sonnenberg, A. (1990) *EMBO J.* **9,** 765–770
19. Suzuki, S., and Naitoh, Y. (1990) *EMBO J.* **9,** 757–763
20. Tamura, R. N., Rozzo, C., Starr, L., Chambers, J., Reichardt, L. F., Cooper, H. M., and Quaranta, V. (1990) *J. Cell Biol.* **111,** 1593–1604
21. Ginsberg, M. H., Loftus, J. C., Ryckwaert, J.-J., Pierschbacher, M. D., Pytela, R., Ruoslahti, E., and Plow, E. F. (1987) *J. Biol. Chem.* **262,** 5437–5440
22. Hynes, R. O. (1987) *Cell* **48,** 549–554
23. Takada, Y., and Hemler, M. E. (1989) *J. Cell Biol.* **109,** 397–407
24. Cheresh, D. A., Smith, J. W., Cooper, H. M., and Quaranta, V. (1989) *Cell* **57,** 59–69
25. Holzmann, B., and Weissman, I. L. (1989) *EMBO J.* **8,** 1735–1741
26. Kajiji, S., Tamura, R. N., and Quaranta, V. (1989) *EMBO J.* **8,** 673–680
27. Hemler, M. E., Crouse, C., and Sonnenberg, A. (1989) *J. Biol. Chem.* **264,** 6529–6535
28. Nurden, A. T., and Caen, J. P. (1974) *Br. J. Haematol.* **28,** 253–260
29. Anderson, D. C., Schmalstieg, F. C., Arnaout, M. A., Kohl, S., Tosi, M. F., Dana, N., Buffone, G. J., Hughes, B. J., Brinkley, B. R., Dickey, W. D., Abramson, J. S., Springer, T. A., Boxer, L. A., Hollers, J. M., and Smith, C. W. (1984) *J. Clin. Invest.* **74,** 536–551
30. Springer, T. A., Thompson, W. S., Miller, L. J., Schmalstieg, F. C., and Anderson, D. C. (1984) *J. Exp. Med.* **160,** 1901–1918
31. Wilcox, M., Brown, N., Piovant, M., Smith, R. J., and White, A. H. (1984) *EMBO J.* **3,** 2307–2313
32. MacKrell, A. J., Blumberg, B., Haynes, S. R., and Fessler, J. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2633–2637
33. Yang, J. T., Rayburn, H., and Hynes, R. O. (1993) *Development (Camb.)* **119,** 1093–1105
34. Ginsberg, M. H., Du, X., and Plow, E. F. (1992) *Curr. Opin. Cell Biol.* **4,** 766–771
35. Schwartz, M. A. (1992) *Trends Cell Biol.* **2,** 304–308
36. Singer, S. J. (1992) *Science* **255,** 1671–1677
37. Juliano, R. (1992) *Cancer Metastasis Rev.* **13,** 25–30
38. Ginsberg, M. H., Forsyth, J., Lightsey, A., Chediak, J., and Plow, E. F. (1983) *J. Clin. Invest.* **71,** 619–624
39. Pytela, R., Pierschbacher, M. D., and Ruoslahti, E. (1985) *Cell* **40,** 191–198
40. Parise, L. V., and Phillips, D. R. (1986) *J. Biol. Chem.* **261,** 14011–14017
41. Marlin, S. D., and Springer, T. A. (1987) *Cell* **51,** 813–819
42. Isberg, R. R., and Leong, J. M. (1990) *Cell* **60,** 861–871
43. Relman, D., Tuomanen, E., Falkow, S., Golenbock, D. T., Saukkonen, K., and Wright, S. D. (1990) *Cell* **61,** 1375–1382
44. Fernandez, M. A., Munoz-Fernandez, M. A., and Fresno, M. (1993) *Eur. J. Immunol.* **23,** 552–557
45. Bennett, J. S., and Vilaire, G. (1979) *J. Clin. Invest.* **64,** 1393–1401
46. Bennett, J. S., Vilaire, G., and Cines, D. B. (1982) *J. Biol. Chem.* **257,** 8049–8054
47. Ginsberg, M. H., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. A., and Plow, E. F. (1985) *J. Biol. Chem.* **260,** 3931–3936
48. Pierschbacher, M. D., and Ruoslahti, E. (1984) *Nature* **309,** 30–33
49. Pierschbacher, M. D., and Ruoslahti, E. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 5985–5988
50. Ruggeri, Z. M., Houghten, R. A., Russell, S. R., and Zimmerman, T. S. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 5708–5712
51. Humphries, M. J., Akiyama, S. K., Komoriya, A., Olden, K., and Yamada, K. M. (1986) *J. Cell Biol.* **103,** 2637–2647
52. Komoriya, A., Green, L. J., Mervic, M., Yamada, S. S., Yamada, K. M., and Humphries, M. J. (1991) *J. Biol. Chem.* **266,** 15075–15079
53. Vonderheide, R. H., Tedder, T. F., Springer, T. A., and Staunton, D. E. (1994) *J. Cell Biol.* **125,** 215–222
54. Osborn, L., Vassallo, C., Browning, B. G., Tizard, R., Haskard, D. O., Benjamin, C. D., Dougas, I., and Kirchhausen, T. (1994) *J. Cell Biol.* **124,** 601–608
55. Berendt, A. R., McDowall, A., Craig, A. G., Bates, P. A., Sternberg, M. J. E., Marsh, K., Newbold, C. I., and Hogg, N. (1992) *Cell* **68,** 71–81
56. Obara, M., Kang, M. S., and Yamada, K. M. (1988) *Cell* **53,** 649–657
57. Bowditch, R. D., Halloran, C. E., Aota, S., Obara, M., Plow, E. F., Yamada, K. M., and Ginsberg, M. H. (1991) *J. Biol. Chem.* **266,** 23323–23328
58. Bowditch, R. D., Hariharan, M., Tominna, E. F., Smith, J. W., Yamada, K. M., Getzoff, E. D., and Ginsberg, M. H. (1994) *J. Biol. Chem.* **269,** 10856–10863
59. Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992) *Science* **258,** 987–991
60. Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) *Cell* **71,** 671–678
61. Adler, M., Lazarus, R. A., Dennis, M. S., and Wagner, G. (1991) *Science* **253,** 445–448
62. Taub, R., Gould, R. J., Garsky, V. M., Ciccarone, T. M., Hoxie, J., Friedman, P. A., and Shattil, S. J. (1989) *J. Biol. Chem.* **264,** 259–265
63. Tomiyama, Y., Brojer, E., Ruggeri, Z. M., Shattil, S. J., Smiltneck, J., Gorski, J., Kumar, A., Kieber-Emmons, T., and Kunicki, T. J. (1992) *J. Biol. Chem.* **267,** 18085–18092
64. Nermut, M. V., Green, N. M., Eason, P., Yamada, S. S., and Yamada, K. M. (1988) *EMBO J.* **7,** 4093–4099
65. Carrell, N. A., Fitzgerald, L. A., Steiner, B., Erickson, H. P., and Phillips, D. R. (1985) *J. Biol. Chem.* **260,** 1743–1749
66. Weisel, J. W., Nagaswami, C., Vilaire, G., and Bennett, J. S. (1992) *J. Biol. Chem.* **267,** 16637–16643
67. Du, X., Gu, M., Weisel, J. W., Nagaswami, C., Bennett, J. S., Bowditch, R. D., and Ginsberg, M. H. (1993) *J. Biol. Chem.* **268,** 23087–23092
68. Parise, L. V., and Phillips, D. R. (1985) *J. Biol. Chem.* **260,** 1750–1756
69. Jennings, L. K., and Phillips, D. R. (1982) *J. Biol. Chem.* **257,** 10458–10466
70. Hantgan, R. R., Braaten, J. V., and Rocco, M. (1993) *Biochemistry* **32,** 3935–3941
71. Calvete, J. J., Henschen, A., and Gonzalez-Rodriguez, J. (1991) *Biochem. J.* **274,** 63–71

72. Calvete, J. J., Henschen, A., and Gonzalez-Rodriguez, J. (1989) *Biochem. J.* **261**, 561–568
73. Calvete, J. J., Arias, J., Alvarez, M. V., Lopez, M. M., Henschen, A., and Gonzalez-Rodriguez, J. (1991) *Biochem. J.* **273**, 767–775
74. Calvete, J. J., Alvarez, M. V., Rivas, G., Hew, C. L., Henschen, A., and Gonzalez-Rodriguez, J. (1989) *Biochem. J.* **261**, 551–560
75. Calvete, J. J., Mann, K., Alvarez, M. V., Lopez, M. M., and Gonzalez-Rodriguez, J. (1992) *Biochem. J.* **282**, 523–532
76. Calvete, J. J., Mann, K., Schafer, W., Fernandez-Lafuente, R., and Guisan, J. M. (1994) *Biochem. J.* **298**, 1–7
77. Rocco, M., Spotorno, B., and Hantgan, R. R. (1993) *Protein Sci.* **2**, 2154–2166
78. Parise, L. V., Helgerson, S. L., Steiner, B., Nannizzi, L., and Phillips, D. R. (1987) *J. Biol. Chem.* **262**, 12597–12602
79. Steiner, B., Parise, L. V., Leung, B., and Phillips, D. R. (1991) *J. Biol. Chem.* **266**, 14986–14991
80. Frelinger, A. L., III, Lam, S. C.-T., Plow, E. F., Smith, M. A., Loftus, J. C., and Ginsberg, M. H. (1988) *J. Biol. Chem.* **263**, 12397–12402
81. Frelinger, A. L., III, Cohen, I., Plow, E. F., Smith, M. A., Roberts, J., Lam, S. C.-T., and Ginsberg, M. H. (1990) *J. Biol. Chem.* **265**, 6346–6352
82. Pidard, D., Frelinger, A. L., III, Bouillot, C., and Nurden, A. T. (1991) *Eur. J. Biochem.* **200**, 437–447
83. Phillips, D. R., and Agin, P. P. (1977) *J. Biol. Chem.* **252**, 2121–2126
84. Fitzgerald, L. A., and Phillips, D. R. (1985) *J. Biol. Chem.* **260**, 11366–11374
85. Rivas, G. A., Calvete, J. J., and Gonzalez-Rodriguez, J. (1991) *Protein Expression Purif.* **2**, 248–255
86. Lam, S. C.-T. (1992) *J. Biol. Chem.* **267**, 5649–5655
87. Wippler, J., Kouns, W. C., Schlaeger, E. J., Kuhn, H., Hadvary, P., and Steiner, B. (1994) *J. Biol. Chem.* **269**, 8754–8761
88. Santoro, S. A., and Lawing, W. J., Jr. (1987) *Cell* **48**, 867–873
89. Smith, J. W., and Cheresh, D. A. (1990) *J. Biol. Chem.* **265**, 2168–2172
90. Smith, J. W., and Cheresh, D. A. (1988) *J. Biol. Chem.* **263**, 18726–18731
91. D'Souza, S. E., Ginsberg, M. H., Burke, T. A., Lam, S. C.-T., and Plow, E. F. (1988) *Science* **242**, 91–93
92. Pidard, D., Didry, D., Kunicki, T. J., and Nurden, A. T. (1986) *Blood* **67**, 604–611
93. Brass, L. F., Shattil, S. J., Kunicki, T. J., and Bennett, J. S. (1985) *J. Biol. Chem.* **260**, 7875–7881
94. Kunicki, T. J., Pidard, D., Rosa, J., and Nurden, A. T. (1981) *Blood* **58**, 268–277
95. Shattil, S. J., Brass, L. F., Bennett, J. S., and Pandhi, P. (1985) *Blood* **66**, 92–98
96. Buck, C. A., Shea, E., Duggan, K., and Horwitz, A. F. (1986) *J. Cell Biol.* **103**, 2421–2428
97. Loftus, J. C., O'Toole, T. E., Plow, E. F., Glass, A., Frelinger, A. L., III, and Ginsberg, M. H. (1990) *Science* **249**, 915–918
98. Bajt, M. L., and Loftus, J. C. (1994) *J. Biol. Chem.* **269**, 20913–20919
99. Andrieux, A., Rabiet, M. J., Chapel, A., Concord, E., and Marguerie, G. (1991) *J. Biol. Chem.* **266**, 14202–14207
100. Calvete, J. J., Arias, J., Alvarez, M. V., Lopez, M. M., Henschen, A., and Gonzalez-Rodriguez, J. (1991) *Biochem. J.* **274**, 457–463
101. Wang, R., Furihata, K., McFarland, J. G., Friedman, K., Aster, R. H., and Newman, P. J. (1992) *J. Clin. Invest.* **90**, 2038–2043
102. Bajt, M. L., Loftus, J. C., Gawaz, M. P., and Ginsberg, M. H. (1992) *J. Biol. Chem.* **267**, 22211–22216
103. Takada, Y., Ylanne, J., Mandelman, D., Puzon, W., and Ginsberg, M. H. (1992) *J. Cell Biol.* **119**, 913–921
104. Huang, X., Chen, A., Agrez, M., and Sheppard, D. (1994) *Mol. Biol. Cell* **4**, 282a (abstr.)
105. Ginsberg, M. H., Lightsey, A., Kunicki, T. J., Kaufman, A., Marguerie, G. A., and Plow, E. F. (1986) *J. Clin. Invest.* **78**, 1103–1111
106. Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem.* **8**, 119–174
107. Haas, T. A., D'Souza, S. E., Cierniewski, C. S., and Plow, E. F. (1993) *Thromb. Haemostasis* **69**, 786a (abstr.)
108. D'Souza, S. E., Haas, T. A., Piotrowicz, R. S., Byers-Ward, V., McGrath, D. E., Soule, H. R., Cierniewski, C., Plow, E. F., and Smith, J. W. (1994) *Cell*, in press
109. Charo, I. F., Nannizzi, L., Phillips, D. R., Hsu, M. A., and Scarborough, R. M. (1991) *J. Biol. Chem.* **266**, 1414–1421
110. Steiner, B., Trzeciak, A., Pfenninger, G., and Kouns, W. C. (1993) *J. Biol. Chem.* **268**, 6870–6873
111. Bajt, M. L., Ginsberg, M. H., Frelinger, A. L., III, Berndt, M. C., and Loftus, J. C. (1992) *J. Biol. Chem.* **267**, 3789–3794
112. Lanza, F., Stierle, A., Fournier, D., Morales, M., Andre, G., Nurden, A. T., and Cazenave, J.-P. (1992) *J. Clin. Invest.* **89**, 1995–2004
113. Fournier, D. J., Kabral, A., Castaldi, P. A., and Berndt, M. C. (1989) *Thromb. Haemostasis* **62**, 977–983
114. D'Souza, S. E., Ginsberg, M. H., Burke, T. A., and Plow, E. F. (1990) *J. Biol. Chem.* **265**, 3440–3446
115. D'Souza, S. E., Ginsberg, M. H., Matsueda, G. R., and Plow, E. F. (1991) *Nature* **350**, 66–68
116. Taylor, D. B., and Gartner, T. K. (1992) *J. Biol. Chem.* **267**, 11729–11733
117. Gulino, D., Boudignon, C., Zhang, L. Y., Concord, E., Rabiet, M. J., and Marguerie, G. (1992) *J. Biol. Chem.* **267**, 1001–1007
118. Wilcox, D. A., Wautier, J. L., Pidard, D., and Newman, P. J. (1994) *J. Biol. Chem.* **269**, 4450–4457
119. Stanley, P., Bates, P. A., Harvey, J., Bennett, R. I., and Hogg, N. (1994) *EMBO J.* **13**, 1790–1798
120. Hemler, M. E. (1990) *Annu. Rev. Immunol.* **8**, 365–400
121. Shaw, S. K., Cepek, K. L., Murphy, E. A., Russell, G. J., Brenner, M. B., and Parker, C. M. (1994) *J. Biol. Chem.* **269**, 6016–6025
122. Fleming, J. C., Pahl, H. L., Gonzalez, D. A., Smith, T. F., and Tenen, D. G. (1993) *J. Immunol.* **150**, 480–490
123. Hughes, A. L. (1992) *Mol. Biol. Evol.* **9**, 216–234
124. Larson, R. S., Corbi, A. L., Berman, L., and Springer, T. A. (1989) *J. Cell Biol.* **108**, 703–712
125. Springer, T. A. (1990) *Nature* **346**, 425–434
126. Bonthron, D., Orr, E. C., Mitsock, L. M., Ginsburg, D., Handin, R. I., and Orkin, S. H. (1986) *Nucleic Acids Res.* **14**, 7125–7127
127. Shelton-Inloes, B. B., Titani, K., and Sadler, J. E. (1986) *Biochemistry* **25**, 3164–3171
128. Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., and Fujikawa, K. (1986) *Biochemistry* **25**, 3171–3184
129. Verweij, C. L., Diergaarde, P. J., Hart, M., and Pannekoek, H. (1986) *EMBO J.* **5**, 1839–1847
130. Corbi, A. L., Kishimoto, T. K., Miller, L. J., and Springer, T. A. (1988) *J. Biol. Chem.* **263**, 12403–12411
131. Diamond, M. S., Garcia-Aguilar, J., Bickford, J. K., Corbi, A. L., and Springer, T. A. (1993) *J. Cell Biol.* **120**, 1031–1043
132. Michishita, M., Videm, V., and Arnaout, M. A. (1993) *Cell* **72**, 857–867
133. Kamata, T., Puzon, W., and Takada, Y. (1994) *J. Biol. Chem.* **269**, 9659–9663
134. Kiss, I., Deak, F., Holloway, R. G. J., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., and Goetinck, P. F. (1989) *J. Biol. Chem.* **264**, 8126–8134
135. Bonaldo, P., Russo, V., Bucciotti, F., Bressan, G. M., and Colombatti, A. (1989) *J. Biol. Chem.* **264**, 5575–5580
136. Robson, K. J., Hall, J. R., Jennings, M. W., Harris, T. J., Marsh, K., Newbold, C. I., Tate, V. E., and Weatherall, D. J. (1988) *Nature* **335**, 79–82
137. Dransfield, I., Cabanas, C., Craig, A., and Hogg, N. (1992) *J. Cell Biol.* **116**, 219–226
138. Staatz, W. D., Rajpara, S. M., Wayner, E. A., Carter, W. G., and Santoro, S. A. (1989) *J. Cell Biol.* **108**, 1917–1924
139. Staatz, W. D., Peters, K. J., and Santoro, S. A. (1990) *Biochem. Biophys. Res. Commun.* **168**, 107–113
140. Shadle, P. J., and Barondes, S. H. (1982) *J. Cell Biol.* **95**, 361–365
141. Li, R., Nortamo, P., Kantor, C., Kovanen, P., Timonen, T., and Gahmberg, C. G. (1993) *J. Biol. Chem.* **268**, 21474–21477
142. Ross, L., Hassman, F., and Molony, L. (1992) *J. Biol. Chem.* **267**, 8537–8543
143. Staatz, W. D., Fok, K. F., Zutter, M. M., Adams, S. P., Rodriguez, B. A., and Santoro, S. A. (1991) *J. Biol. Chem.* **266**, 7363–7367
144. Cheng, S., Craig, W. S., Mullen, D., Tschopp, J. F., Dixon, D., and Pierschbacher, M. D. (1994) *J. Med. Chem.* **37**, 1–8
145. Barbas, C. F., Languino, L. R., and Smith, J. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10003–10007
146. Greenspoon, N., Hershkoviz, R., Alon, R., Varon, D., Shenkman, B., Marx, G., Federman, S., Kapustina, G., and Lider, O. (1993) *Biochemistry* **32**, 1001–1008
147. Horton, M. A., Dorey, E. L., Nesbitt, S. A., Samanen, J., Ali, F. E., Stadel, J. M., Nichols, A., Greig, R., and Helfrich, M. H. (1993) *J. Bone Miner. Res.* **8**, 239–247
148. Koivunen, E., Gay, D. A., and Ruoslahti, E. (1993) *J. Biol. Chem.* **268**, 20205–20210
149. Nowlin, D. M., Gorcsan, F., Moscinski, M., Chiang, S. L., Lobl, T. J., and Cardarelli, P. M. (1993) *J. Biol. Chem.* **268**, 20352–20359
150. Alig, L., Edenhofer, A., Hadvary, P., Hurzeler, M., Knopp, D., Muller, M., Steiner, B., Trzeciak, A., and Weller, T. (1992) *J. Med. Chem.* **35**, 4393–4407
151. Strynadka, N. C. J., and James, M. N. G. (1990) *Proteins Struct. Funct. Genet.* **7**, 234–248
152. Tuckwell, D. S., Brass, A., and Humphries, M. J. (1992) *Biochem. J.* **285**, 325–331
153. Lawler, J., and Hynes, R. O. (1986) *J. Cell Biol.* **103**, 1635–1648
154. Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S., and Springer, T. A. (1987) *EMBO J.* **6**, 4023–4028
155. Smith, J. W., and Cheresh, D. A. (1991) *J. Biol. Chem.* **266**, 11429–11432
156. Martin-Padura, I., Bazzoni, G., Zanetti, A., Bernasconi, S., Elices, M. J., Mantovani, A., and Dejana, E. (1994) *J. Biol. Chem.* **269**, 6124–6132
157. Miles, L. A., Ginsberg, M. H., White, J. G., and Plow, E. F. (1986) *J. Clin. Invest.* **77**, 2001–2009
158. Myohanen, H. T., Stephens, R. W., Hedman, K., Tapiovaara, H., Ronne, E., Hoyer-Hansen, G., Dano, K., and Vaheri, A. (1993) *J. Histochem. Cytochem.* **41**, 1291–1301
159. Loftus, J. C., Plow, E. F., Jennings, L., and Ginsberg, M. H. (1988) *J. Biol. Chem.* **263**, 11025–11028
160. Calvete, J. J., Schafer, W., Henschen, A., and Gonzalez-Rodriguez, J. (1990) *FEBS Lett.* **27**, 237–240
161. Larjava, H., Peltonen, J., Akiyama, S. K., Yamada, S. S., Gralnick, H. R., Uitto, J., and Yamada, K. M. (1990) *J. Cell Biol.* **110**, 803–815
162. Carter, W. G., Wayner, E. A., Bouchard, T. S., and Kaur, P. (1990) *J. Cell Biol.* **110**, 1387–1404
163. Lampugnani, M. G., Resnati, M., Dejana, E., and Marchisio, P. C. (1991) *J. Cell Biol.* **112**, 479–490
164. Sriramarao, P., Steffner, P., and Gehlsen, K. R. (1993) *J. Biol. Chem.* **268**, 22036–22041
165. Qian, F., Vaux, D. L., and Weissman, I. L. (1994) *Cell* **77**, 335–347
166. Campanero, M. R., del Pozo, M. A., Arroyo, A. G., Sanchez-Mateos, P., Hernandez-Caselles, T., Craig, A., Pulido, R., and Sanchez-Madrid, F. (1993) *J. Cell Biol.* **123**, 1007–1016
167. Pulido, R., Elices, M. J., Campanero, M. R., Osborn, L., Schiffer, S., Garcia-Pardo, A., Lobb, R., Hemler, M. E., and Sanchez-Madrid, F. (1991) *J. Biol. Chem.* **266**, 10241–10245
168. Campanero, M. R., Pulido, R., Ursa, M. A., Rodriguez Moya, M., de Landazuri, M. O., and Sanchez-Madrid, F. (1990) *J. Cell Biol.* **110**, 2157–2165
169. Bergelson, J. M., Shepley, M. P., Chan, B. M. C., Hemler, M. E., and Finberg, R. W. (1992) *Science* **255**, 1718–1720
170. Deleted in proof

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 357–361, January 1996
Medical Sciences

Cloning and expression of a gene encoding an integrin-like protein in *Candida albicans*

CHERYL GALE*, DAVID FINKEL*, NIANJUN TAO*, MARILYN MEINKE*, MARK MCCLELLAN*, JENNIFER OLSON*, KATHLEEN KENDRICK†, AND MARGARET HOSTETTER*‡

*Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455; and †Department of Microbiology, Ohio State University, Columbus, OH 43210

Communicated by Mary Ellen Avery, Harvard Medical School, Boston, MA, September 26, 1995

ABSTRACT **The existence of integrin-like proteins in *Candida albicans* has been postulated because monoclonal antibodies to the leukocyte integrins αM and αX bind to blastospores and germ tubes, recognize a candidal surface protein of ≈185 kDa, and inhibit candidal adhesion to human epithelium. The gene *αINT1* was isolated from a library of *C. albicans* genomic DNA by screening with a cDNA probe from the transmembrane domain of human αM. The predicted polypeptide (αInt1p) of 188 kDa contains several motifs common to αM and αX: a putative I domain, two EF-hand divalent cation-binding sites, a transmembrane domain, and a cytoplasmic tail with a single tyrosine residue. An internal RGD tripeptide is also present. Binding of anti-peptide antibodies raised to potential extracellular domains of αInt1p confirms surface localization in *C. albicans* blastospores. By Southern blotting, *αINT1* is unique to *C. albicans*. Expression of *αINT1* under control of a galactose-inducible promoter led to the production of germ tubes in haploid *Saccharomyces cerevisiae* and in the corresponding *ste12* mutant. Germ tubes were not observed in haploid yeast transformed with vector alone, in transformants expressing a galactose-inducible gene from *Chlamydomonas*, or in transformants grown in the presence of glucose or raffinose. Transformants producing αInt1p bound an anti-αM monoclonal antibody and exhibited enhanced aggregation. Studies of αInt1p reveal novel roles for primitive integrin-like proteins in adhesion and in *STE12*-independent morphogenesis.**

The opportunistic pathogen *Candida albicans* is the leading cause of invasive fungal disease in neonates, diabetics, and immunocompromised patients and carries a high mortality despite prompt and appropriate anti-fungal therapy (1–3). Three important events in the pathogenesis of invasive candidal infection include adhesion to epithelium, penetration of epithelial barriers, and hematogenous dissemination. Complicating this cascade is the yeast's ability to transform from blastospores at the epithelial surface to elongated structures (germ tubes, pseudohyphae, and mycelia) that invade underlying tissues.

Several investigators have reported the existence of surface proteins in *C. albicans* that are antigenically, structurally, and functionally related to the α-subunits of the leukocyte integrins αM/β2 (Mac-1; CD11b/CD18) and αX/β2 (p150,95; CD11c/CD18) (4–11). Many monoclonal antibodies (mAbs) recognizing epitopes of αM or αX bind to blastospores or germ tubes of *C. albicans* (4–10). iC3b-coated sheep erythrocytes rosette with germ tubes of *C. albicans* (9), and the affinity constants for the binding of purified human iC3b to *C. albicans* or to leukocyte αM/β2 are virtually identical (5, 12). Environmental conditions such as increased temperature or glucose concentrations ≥10 mM augment not only the surface expression of this integrin-like protein (5, 11) but also the binding of iC3b (12). Recognition of ligands containing the tripeptide sequence arginine-glycine-aspartic acid (RGD) facilitates the adhesion of *C. albicans* to endothelial and epithelial cells (6, 11).

Among the leukocyte integrins, αM and αX share ≈70% sequence homology and considerable functional identity. These two α-subunits, together with αL, α1, and α2, contain an inserted or I domain of ≈200 amino acids that is involved in ligand binding (13–15). Located just C-terminal to the I domain in αM/αX are three divalent cation-binding sites; at the C terminus are a membrane-spanning region and a cytoplasmic tail, the latter containing a single tyrosine residue in αM and αX (13).

This manuscript reports the isolation of a *C. albicans* gene encoding a protein that shares these integrin motifs.§ Moreover, expression of the gene product in haploid *Saccharomyces cerevisiae* is associated with the production of germ tubes independently of Ste12p, a yeast transcription factor required for morphologic change in response to mating pheromones and nutrient limitations in *S. cerevisiae* (16). These results open the way for the discovery of other integrin-like proteins in primitive eukaryotes, for their study as precursors of vertebrate integrins, and for more detailed investigation of their roles in signal transduction and morphogenesis.

MATERIALS AND METHODS

Yeast Strains, Plasmids, and Reagents for Cloning. *C. albicans* 10261 (B311, serotype A) was purchased as a lyophilate (American Type Culture Collection). *Candida tropicalis* 7555 was isolated from the blood of a fungemic patient by the University of Minnesota Clinical Microbiology Laboratory. *S. cerevisiae* YPH500 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1*) is a galactose-utilizing strain obtained from the Yeast Genetic Stock Center (Berkeley, CA) (17). pBM272, an ARS/CEN-based yeast shuttle vector containing the *URA3* gene and the *S. cerevisiae GAL1-10* promoter (18), pGG201 containing a 990-bp open reading frame encoding a DNA-binding protein from *Chlamydomonas reinhardtii* (19), a 750-bp *Cla* I/*Hin*dIII fragment of the *C. albicans* actin gene, and *S. cerevisiae* strain M12B-T2 were gifts from James Bodley, Judith Berman, Paul Magee, and Beatrice Magee (all of the University of Minnesota), respectively. pSUL16, a gift from Judith Berman, contains the *S. cerevisiae STE12* gene interrupted with the yeast *LEU2* gene (20). *Escherichia coli* JM101, LE392, XL1-Blue-MRF', and pBluescript II SK(+) were obtained from Stratagene.

Cloning of *αINT1*. DNA from spheroplasts of *C. albicans* 10261 was isolated according to standard procedures (21), digested with *Sau*3AI, and packaged in λEMBL3 (Stratagene).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MM, minimal medium; mAb, monoclonal antibody.

‡To whom reprint requests should be addressed at: University of Minnesota, Department of Pediatrics, Box 296 UMHC, 420 Delaware Street, S.E., Minneapolis, MN 55455.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. U35070).

Preliminary studies confirmed that a 3.5-kbp *Eco*RI fragment of *C. albicans* DNA hybridized with a 314-bp *Eco*RI/*Sma* I cDNA fragment derived from the transmembrane domain of human αM (kind gift of Dennis Hickstein, Veterans' Administration Medical Center, Seattle). A library enriched for 3.0- to 3.8-kbp *Eco*RI fragments was constructed by digestion of genomic DNA with *Eco*RI and ligation to pBluescript II SK(+). Primers for amplification of the *Eco*RI/*Sma* I αM cDNA fragment were as follows: upstream primer, 5'-GAATTCAATGCTACCCTCAA; downstream primer, 5'-CCCGGGGGACCCCCTTCACT. Plasmid minipreparations from a total of 200 colonies were screened by the sib selection technique for hybridization at 50°C with ^{32}P-labeled PCR product after confirmation of its nucleotide sequence (13). Five clones were isolated from three successive screenings. Two of the five clones gave reproducible signals after hybridization with a degenerate oligonucleotide encoding a conserved sequence (KVGFFK) in the cytoplasmic domain of αX (22): 5'-AA(AG) GT(CT) GG(AT) TT(CT) TT(CT) AA(AG)-3'. Both clones contained a 3.5-kbp *Eco*RI insert and failed to hybridize with a degenerate oligonucleotide from the *S. cerevisiae* gene *USO1* (23): 5'-GAA AT(ACT) GA(CT) GA(CT) TT(AG) ATG-3'. One of these clones (probe 2, Fig. 1*A*) was chosen for further analysis. A 500-bp *Hin*dIII subfragment (probe 3, Fig. 1*A*) was used to screen 20,000 clones from a library of *C. albicans*

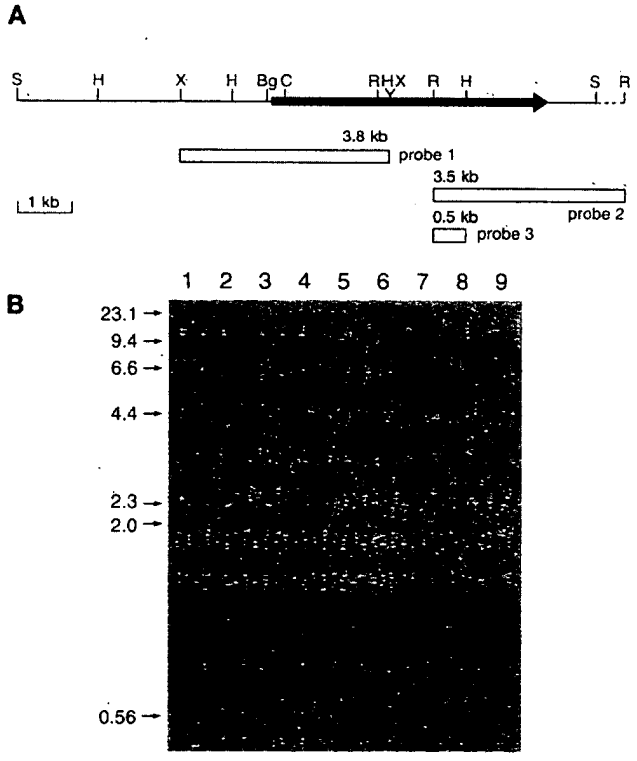


FIG. 1. (*A*) Restriction map of the 10.5-kbp *Sal* I genomic DNA fragment isolated from *C. albicans* 10261, with the open reading frame indicated by the bold arrow. Probe 1, 3.8-kbp *Xba* I fragment used for Southern and Northern blotting. S, *Sal* I; H, *Hin*dIII; X, *Xba* I; Bg, *Bgl* II; C, *Cla* I; R, *Eco*RI. (*B*) Southern blot of genomic DNA from *C. albicans* 10261 (lanes 1, 4, and 7), *C. tropicalis* 7555 (lanes 2, 5, and 8), and *S. cerevisiae* YPH500 (lanes 3, 6, and 9) digested with *Eco*RI (lanes 1–3), *Hin*dIII (lanes 4–6), and *Xba* I (lanes 7–9) and hybridized at high stringency with [α-^{32}P]dGTP-labeled probe 1 (hybridization at 65°C, final wash in 0.2× SSC/0.1% SDS at 65°C). The high molecular weight band (>12 kbp in lane 7) most likely represents incompletely digested DNA. Positions of *Hin*dIII-digested λDNA fragments are indicated on the far left. *Eco*RI and *Hin*dIII digests of four additional *S. cerevisiae* isolates from clinical and laboratory sources, as well as isolates of *Candida glabrata* and *Candida parapsilosis*, also failed to hybridize with probe 1.

10261 genomic DNA by the plaque hybridization technique (24). The largest hybridizing insert, a 10.5-kbp *Sal* I fragment (Fig. 1*A*), was isolated by agarose gel electrophoresis, cloned, and sequenced.

Sequence Analysis. Both strands of the 10.5-kbp *Sal* I fragment were sequenced by the method of gene walking on an Applied Biosystems model 373 automated sequencer in the University of Minnesota Microchemical Facility. Nucleotide and protein sequence analyses were performed with the Genetics Computer Group (University of Wisconsin, Madison) Sequence Analysis Software Package, version 7.0 (25).

Yeast Transformation and Gene Expression. The entire open reading frame of *αINT1* (*Bgl* II/*Sal* I fragment) was subcloned into pBM272 after digestion with *Bam*HI and *Sal* I, in order to place the *GAL1-10* promoter upstream of the *αINT1* start codon (pCG01). *S. cerevisiae* YPH500 was transformed with pBM272 or pCG01 by the lithium acetate procedure (26). Transformants were selected on agar-based minimal medium (MM = 0.17% yeast nitrogen base/0.5% ammonium sulfate) with 2% glucose, in the absence of uracil. Induction of *αINT1* was achieved by growing transformants containing pCG01 to mid-exponential phase in noninducing, nonrepressing medium (MM without uracil with 2% raffinose) at 30°C, then harvesting, washing, and resuspending them in inducing medium (MM without uracil with 2% galactose) at 30°C. YPH500 and YPH500 transformed with vector alone (pBM272) were grown under the identical conditions.

Southern and Northern Blotting. Genomic DNA and total RNA were isolated and electrophoresed by standard methods (27–30) and transferred to Hybond N+ nylon membranes (Amersham) by traditional capillary blotting.

Flow Cytometry. Anti-peptide antibodies were prepared in rabbits (Cocalico Biologicals, Reamstown, PA) to a 23-mer peptide encompassing the second divalent cation-binding site [amino acid (aa) 596–618] and to a 17-mer peptide spanning the RGD site and flanking residues (aa 1142–1158) in *αINT1*. The IgG fractions of preimmune and immune rabbit sera were isolated on protein A-Sepharose (Pharmacia). Fluorescein isothiocyanate (FITC)-conjugated goat anti-rabbit IgG (Southern Biotechnology Associates) was used as the secondary antibody. For experiments with *S. cerevisiae* transformants, antibodies included OKM1 (anti-αM IgG2b) or MY9 as isotype control (Coulter) and FITC-conjugated goat anti-mouse IgM/IgG (Biosource International, Camarillo, CA) (7, 11).

Insertional Inactivation of *STE12* in *S. cerevisiae*. YPH500 was transformed with pSUL16 by standard techniques (26) and chromosomal integrants of the disrupted *STE12* gene were selected on leucine-deficient MM. After confirmation of sterility, *ste12* mutants were transformed with pCG01 as described above.

Aggregation Assay. The degree of aggregation of *C. albicans* and *S. cerevisiae* transformants was determined according to published methods (31).

RESULTS

Restriction Map and Southern Blotting. The restriction map of *αINT1* with its 5' and 3' flanking sequences is displayed in Fig. 1*A*. Fig. 1*B* shows that a 3.8-kbp *Xba* I probe from *αINT1* hybridized with *Eco*RI, *Hin*dIII, and *Xba* I fragments from *C. albicans* (lanes 1, 4, and 7) but not from *C. tropicalis* 7555 or *S. cerevisiae* YPH500. Among the yeast strains tested, this DNA fragment is unique to *C. albicans*.

Sequence Analysis of *αINT1*. Analysis of the nucleotide sequence revealed an open reading frame sufficient to encode a 1664-residue polypeptide with a theoretical molecular mass of 187,989 Da and no extensive homologies with other proteins. Fig. 2 compares the derived aa sequence of αInt1p with the characteristic motifs of several integrin α-subunits. BESTFIT

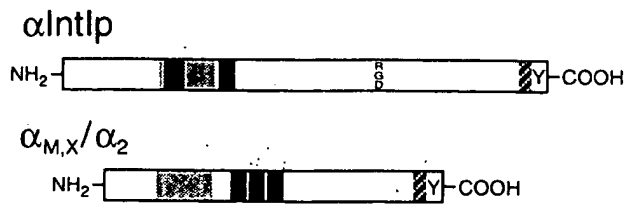


FIG. 2. Schematic diagram comparing the structures of αM, αX, and α2 with that of αIntlp. Gray regions represent the ligand-binding or I domain, the EF-hand divalent cation-binding motifs are indicated in black, and the transmembrane regions are hatched. RGD indicates the approximate location of this sequence in αIntlp (aa 1149–1151). The α-subunit schematic is modified from the sequence reported by Corbi *et al.* (13).

analysis (32) located a putative I domain at aa 230–470, with ≈18% identity to the I domain of human αM. Within this I-domain-like region are three potential partial MIDAS motifs (DXSX) for the coordination of divalent cations (33). This same region (aa 230–470) also displayed 25% identity to the nonrepeat region of the fibrinogen-binding protein of *Staphylococcus aureus* (34). Chou–Fasman analysis (35) indicated multiple α-helices, two of them bracketing the second of two possible EF-hand divalent cation-binding motifs (aa 283–295 and aa 601–613). Fig. 3 shows that the amino acid sequence of the second divalent cation-binding site from αIntlp differs from the EF-hand consensus sequence (36) at only one residue, a non-cation coordinating site. A hydrophobic sequence is located at aa 1592–1617 as determined by Kyte–Doolittle hydrophobicity plotting (37). Just C-terminal to this putative membrane-spanning region in αIntlp is a unique tyrosine residue, also present in the cytoplasmic tails of αM and αX (13, 22).

In the upstream sequence, a putative TATA box is located at −34 from the start codon. The coding sequence also displays 24 N-glycosylation sites, 6 cysteine residues, and the tripeptide sequence arginine-glycine-aspartic acid (RGD) (aa 1149–1151), a feature of many integrin ligands but not of integrins themselves.

Localization of αIntlp in *C. albicans* and *S. cerevisiae*. Polyclonal antibodies prepared against the second divalent cation-binding site and the RGD sequence and flanking residues in αIntlp recognized 64–82% of *C. albicans* blastospores, while preimmune IgG bound to only 0.5–1% of yeast cells ($P < 0.0001$) (Table 1). These results confirm that αIntlp is a surface protein in *C. albicans* and that the second cation-binding site and the RGD site are in the extracellular region of the polypeptide. In *S. cerevisiae*, the binding of the anti-αM mAb OKM1 was significantly greater in transformants expressing *αINT1* vs. transformants containing vector alone for percent yeasts fluorescing (19.0% vs. 6.2%; $P \leq$ 0.004) and for mean channel fluorescence (181.8 vs. 65.7; $P \leq$ 0.013). These results confirm that αIntlp is surface-borne in *S. cerevisiae* transformants and is recognized by an anti-integrin mAb.

A

1	2	3	4	5	6	7
(D) -	X -	(DNS) -	(ILVFYW) -	(DENSTG) -	(DNQGHKR) -	(GP) -

8	9	10	11	12	13
(ILVMC) -	(DENQTSGCA) -	X -	X -	(DE) -	(ILVMFYW)

B

	1	2	3	4	5	6	7	8	9	10	11	12	13
αIntlp $I_{283-295}$	N	-	N	-	N	-	S	-	K	-	N	-	V	-	S	-	D	-	M	-	D	-	S
αIntlp $II_{601-613}$	D	-	S	-	N	-	D	-	G	-	D	-	R	-	E	-	D	-	N	-	D	-	D	-	I

FIG. 3. Comparison of divalent cation-binding motifs. (*A*) Consensus sequence for the 13-residue EF-hand divalent cation-binding motif (36). (*B*) The N-terminal (I) and more distal cation-binding site (II) in αIntlp. The standard single letter code for aa residues is used. (. . .), Acceptable amino acids; {. . .}, unacceptable amino acids; X, any amino acid. Cation coordinating sites are indicated in boldface type.

Table 1. Percent yeasts fluorescing and mean channel fluorescence of *C. albicans* blastospores after incubation with anti-peptide antibodies

Antibody source	% yeasts fluorescing	Mean channel fluorescence
Control 12	1.0 ± 0.5	67.4 ± 24.6
UMN12	82.4 ± 8.6*	317.0 ± 24.7*
Control 13	0.40 ± 0.36	36.4 ± 9.2
UMN13	64.1 ± 2.3*	266.7 ± 9.2*

Values represent the mean ± SD of three experiments done on different days using different aliquots of *C. albicans* 10261. UMN12 is the antibody to the second divalent cation-binding motif and UMN13 is the antibody to the RGD region of αIntlp. Control 12 and 13 are preimmune IgGs from rabbits prior to immunization with UMN12 and UMN13, respectively. A one-tailed Student's *t* test was used for statistical calculations.
*$P < 0.0001$ vs. control in all comparisons.

Expression of *αINT1* in *C. albicans* and *S. cerevisiae*. Hybridization of probe 1 with total RNA isolated from *C. albicans* blastospores detected message of ≈5.5 kb (Fig. 4*A*). In *S. cerevisiae*, *αINT1* message was detected in pCG01 transformants 6 hr after induction with 2% galactose and continued to be expressed for at least 24 hr (Fig. 4*B*, lanes 1 and 3). As expected, message was not detected in pCG01 transformants grown under conditions of repression (Fig. 4*B*, lanes 2 and 4) or in pBM272 transformants (Fig. 4*B*, lanes 5 and 6).

Coincident with the detection of *αINT1* message, the majority of the pCG01 transformants formed elongated cell projections reminiscent of germ tubes (Fig. 5*A*). These structures continued to be present for 24 hr and could be detected at galactose concentrations ≥ 0.05%. pCG01 transformants exhibited polar budding, typical of haploid organisms, rather

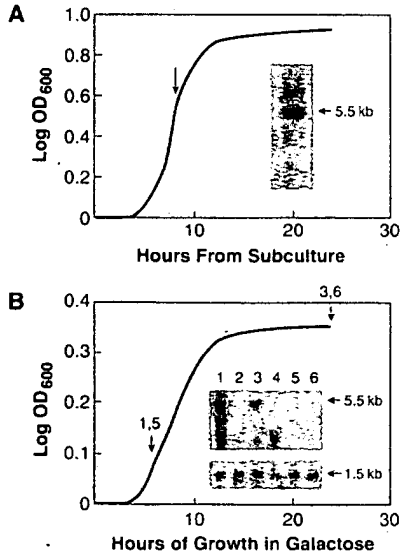


FIG. 4. (*A*) Northern blot of *C. albicans* 10261 total RNA isolated from blastospores in mid-exponential growth (arrow) in MM with 2% glucose and hybridized with probe 1 (see Fig. 1*A*). (*B*) Northern blot of total RNA from *S. cerevisiae* transformants: pCG01 transformants grown in galactose (lanes 1 and 3), pBM272 transformants grown in galactose (lanes 5 and 6), and pCG01 transformants grown to mid-exponential phase (lane 2) and to late exponential phase (lane 4) under conditions of repression (2% glucose). Probe 1 was used for hybridization. The diffuse signal at 2 kbp in lanes 2 and 4 represents nonspecific binding of the probe to the 18S ribosomal RNA band. The signal at 1.5 kbp represents actin transcript.

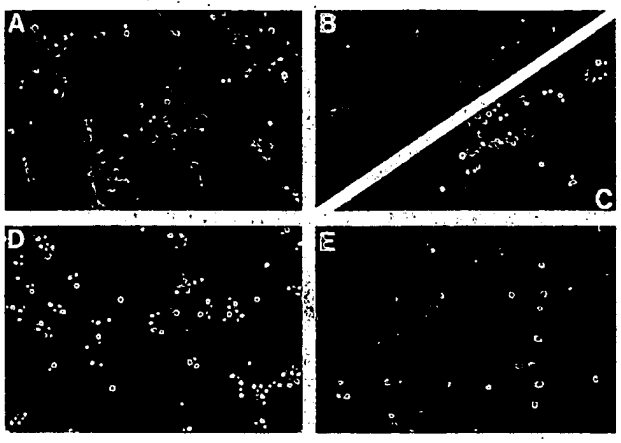


FIG. 5. Phase-contrast photomicrographs of *S. cerevisiae* transformants. pCG01 transformants (*A*) were grown to exponential phase in raffinose and then induced with 2% galactose for 6–24 hr. pBM272 transformants (vector without gene) (*B*), the parent strain YPH500 (*C*), *C. albicans* 10261 (*D*), and pGG201 transformants (galactose-inducible *C. reinhardtii* gene) (*E*) were grown identically. All yeast cells were photographed with a Leitz Wetzlar Laborlux 12 microscope equipped with a WILD MP551 Camera (Heerbrugg, Switzerland). (×500.)

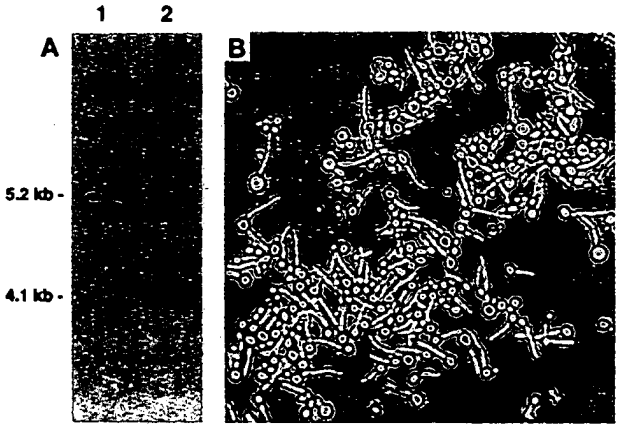


FIG. 6. (*A*) *Cla* I digests of genomic DNA from wild-type YPH500 (lane 1) and *ste12* mutant (lane 2). The blot was probed at high stringency with a 600-bp *Sph* I/*Cla* I fragment from pSUL16. (*B*) Phase-contrast photomicrograph of YPH500 *ste12* mutants transformed with pCG01 and grown in galactose for induction of *αINT1*.

than apical budding, which is typical of diploid organisms. pCG01 transformants (*MATα*) mated to a *MAT***a** yeast strain were able to form diploid organisms (data not shown).

pBM272 transformants, YPH500, and *C. albicans* 10261 did not form germ tubes when grown under the identical conditions (Fig. 5 *B–D*). pCG01 transformants did not exhibit germ tubes when grown in 2% raffinose, 2% glucose, or noninducing concentrations of galactose (0.02%) or when cured of the plasmid (data not shown). In addition, no germ tubes were observed with the galactose-induced expression of an ≈300-residue DNA-binding protein from *C. reinhardtii* (Fig. 5*E*). pCG01 transformants exhibited germ tubes during growth in liquid and on solid medium (MM with 2% galactose). Germ tubes were also observed in yeast strain M12B-T2 transformed with pCG01. Thus, the induction of germ tubes in haploid *S. cerevisiae* is specific to expression of *αINT1* from the plasmid pCG01.

Ability of Yeast Transformants to Aggregate. The aggregation index of pCG01 transformants equaled that of *C. albicans* germ tubes and significantly exceeded the aggregation index of *C. albicans* blastospores and *S. cerevisiae* pBM272 transformants (Table 2). This finding suggests that *S. cerevisiae* germ tubes synthesizing αInt1p are functionally similar to germ tubes in *C. albicans*.

Table 2. Percent aggregation of *C. albicans* and *S. cerevisiae* transformants

Yeast	% aggregation*
C. albicans	
Blastospores	62 ± 1
Hyphae	89 ± 4†
S. cerevisiae	
pBM272	65 ± 4
pCG01	80 ± 2‡

Values represent the mean ± SEM of four experiments, each done in triplicate. *C. albicans* blastospores were grown to mid-exponential phase in YPD medium (1% yeast extract/2% peptone/2% glucose) at 30°C. *C. albicans* hyphae were prepared by growth at 37°C in RPMI 1640 medium (GIBCO/BRL). *S. cerevisiae* pBM272 and pCG01 were grown in galactose-containing medium (see text).

*% aggregation = 100 × (OD_{540} final − OD_{540} initial)/OD_{540} final. A two-tailed Student's *t* test was used to determine statistical significance.

†$P = 0.0013$ vs. *C. albicans* blastospores.

‡$P = 0.026$ vs. *S. cerevisiae* pBM272.

Induction of Germ Tubes in *ste12* Mutants of *S. cerevisiae*. Insertional inactivation of *STE12* in YPH500 shifted the *Cla* I digestion fragment from 5.2 ± 0.1 kbp in the parent to 4.1 ± 0.1 kbp in the *ste12* mutant (Fig. 6*A*). The *Eco*RI fragment shifted from 10.5 ± 0.7 kbp (parent) to 5.0 ± 0.2 kbp (mutant). *ste12* mutants were unable to mate. After transformation of *ste12* mutants with pCG01 and induction of *αINT1* expression by growth in galactose, the mutants formed germ tubes (Fig. 6*B*). Therefore, the observed morphological change is independent of *STE12*.

DISCUSSION

We have isolated a gene encoding a putative integrin-like protein in *C. albicans* by screening a genomic library with conserved sequences from the transmembrane and cytoplasmic domains of human αM. αInt1p exhibits several motifs common to α-integrin subunits, including two EF-hand motifs and three partial MIDAS motifs within a putative I domain, a membrane-spanning domain, and a cytoplasmic tail with a conserved tyrosine residue at the C terminus. Because αM and αX recognize iC3b and fibrinogen as ligands (13), a 25% identity with the fibrinogen-binding protein of *S. aureus* (34) provides additional evidence for relationship.

Divalent cation-binding sites in the amino acid sequence of αM provided initial evidence of the leukocyte integrins' relationship to other vertebrate integrins (13). Both cation-binding motifs in αInt1p conform to the classic EF-hand consensus sequence. In comparison, two of the three cation-binding sites in αM agree at 11 of 13 residues; one of these and the third site require a gap to improve the alignment (13). Chou–Fasman analysis indicates that both divalent cation-binding sites in αInt1p, but not αM, are bracketed by α-helices, a conformation that facilitates cation binding (38). In addition, αInt1p contains three partial MIDAS motifs (DXSX) within the putative I domain. A full or partial MIDAS motif is present in all members of the I domain superfamily (15, 33). Of note, an I-domain-like sequence in *S. cerevisiae* Uso1p binds iC3b and the anti-αM mAb Mn41 (39) but has no divalent cation-binding sites or MIDAS motifs.

The presence of an I domain and an RGD sequence in the extracellular region of αInt1p should contribute to the adhesive capabilities of this protein. For example, an extracellular RGD sequence in the filamentous hemagglutinin of *Bordetella pertussis* facilitates adhesion of the bacterium to eukaryotic

cells (40). Another putative candidal adhesin is encoded by a 3.3-kbp genomic DNA fragment and enables transformed *S. cerevisiae* to adhere to polystyrene or buccal epithelial cells (41). However, its restriction map differs markedly from that of *αINT1*, and the nucleotide sequence has not been published.

In addition to a role as an adhesin, αInt1p leads to the production of germ tubes in haploid *S. cerevisiae* in a process independent of *STE12*. Although the morphological change correlates with expression of the candidal gene product and not with the production of other foreign proteins, we cannot discount the possibilities that αInt1p unnaturally disrupts the cytoskeletal architecture or the growth cycle or that other recognized morphogenic cascades, such as those involving the *CDC* genes (42, 43), may be implicated.

To date, only two genes that participate in morphogenesis in *C. albicans* have been reported. *ACPR*, also called *CPH1*, encodes a protein of 699 aa that is 74% identical to *S. cerevisiae* Ste12p (44, 45). *STE12* is an essential gene in at least two pathways involved in morphogenesis in *S. cerevisiae*: the induction of pseudohyphae in diploid cells on nitrogen-limited medium (46) and the invasive response of haploid cells on rich solid medium (47). Thus, the induction of germ tubes in *S. cerevisiae* transformants expressing *αINT1* after insertional inactivation of *STE12* suggests a novel pathway for integrin-mediated signaling. The second gene, *PHR1*, encodes an ≈580-aa polypeptide essential for pH-dependent morphogenesis in *C. albicans* (48). *ACPR* and *PHR1* encode intracellular regulatory proteins. The isolation of a gene encoding a surface-borne, integrin-like protein in *C. albicans* and its ability to induce morphological variants in haploid *S. cerevisiae* emphasize potential roles for *αINT1* in pathogenesis, signal transduction, and differentiation in *C. albicans* and *S. cerevisiae*.

C.G. is a St. Jude's Children's Research Hospital Fellow sponsored by the Pediatric Scientist Development Program. This research was also supported by funds from the National Institutes of Health (AI25827 and HD7031), the Pediatric AIDS Foundation, and the American Legion Heart Research Foundation to M.H.

1. Butler, K. M. & Baker, C. J. (1988) *Pediatr. Clin. North Am.* **35**, 343–563.
2. Rayfield, E. J., Ault, M. J., Keusch, G. T., Brothers, M. J., Nechemias, C. & Smith, H. (1982) *Am. J. Med.* **72**, 439–450.
3. Kiehn, T. E., Edwards, F. & Armstrong, D. (1980) *Am. J. Clin. Pathol.* **73**, 518–521.
4. Edwards, J. E., Jr., Gaither, T. A., O'Shea, J. J., Rotrosen, D., Lawley, T. J., Wright, S. A., Frank, M. M. & Green, I. (1986) *J. Immunol.* **137**, 3577–3583.
5. Gilmore, B. J., Retsinas, E. M., Lorenz, J. S. & Hostetter, M. K. (1988) *J. Infect. Dis.* **157**, 38–46.
6. Bendel, C. M. & Hostetter, M. K. (1993) *J. Clin. Invest.* **92**, 1840–1849.
7. Bendel, C. M., St. Sauver, J., Carlson, S. & Hostetter, M. K. (1995) *J. Infect. Dis.* **171**, 1760–1763.
8. Mayer, C. L., Diamond, R. D. & Edwards, J. E., Jr. (1990) *Infect. Immun.* **58**, 3765–3769.
9. Heidenreich, F. & Dierich, M. P. (1985) *Infect. Immun.* **50**, 598–600.
10. Hostetter, M. K., Lorenz, J. S., Preus, L. & Kendrick, K. E. (1990) *J. Infect. Dis.* **171**, 761–768.
11. Gustafson, K. S., Vercellotti, G. M., Bendel, C. M. & Hostetter, M. K. (1991) *J. Clin. Invest.* **87**, 1896–1902.
12. Gordon, D. L., Johnson, G. M. & Hostetter, M. K. (1987) *Immunology* **60**, 553–558.
13. Corbi, A. L., Kishimoto, T. K., Miller, L. J. & Springer, T. A. (1988) *J. Biol. Chem.* **25**, 12403–12411.
14. Hynes, R. O. (1992) *Cell* **69**, 11–25.
15. Colombatti, A., Bonaldo, P. & Doliana, R. (1993) *Matrix* **13**, 297–306.
16. Liu, H., Styles, C. A. & Fink, G. R. (1993) *Science* **262**, 1741–1744.
17. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122**, 19–27.
18. Johnston, M. & Davis, R. W. (1984) *Mol. Cell. Biol.* **4**, 1440–1448.
19. Konkel, L. M., Enomoto, S., Chamberlain, E. M., McCune-Zierath, P., Iyadurai, S. J. & Berman, J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5558–5562.
20. Fields, S. & Herskowitz, I. (1987) *Mol. Cell. Biol.* **7**, 3818–3821.
21. Davis, R. W., Thomas, M., Cameron, J., St. John, T. P., Scherer, S. & Padgett, R. A. (1980) *Methods Enzymol.* **65**, 404–411.
22. Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S. & Springer, T. A. (1987) *EMBO J.* **6**, 4023–4028.
23. Nakajima, H., Hirata, M. A., Ogawa, Y., Yonehara, T., Yoda, K. & Yamasaki, M. (1991) *J. Cell Biol.* **113**, 245–260.
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 108–125.
25. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
26. Ito, H., Jukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153**, 173–178.
27. Hoffman, C. S. & Winston, F. (1987) *Gene* **57**, 267–272.
28. Kohrer, K. & Domdey, H. (1991) *Methods Enzymol.* **194**, 398–401.
29. Southern, E. (1979) *Methods Enzymol.* **68**, 152–176.
30. Kaiser, C., Michaelis, S. & Mitchell, A. (1994) *Methods in Yeast Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 151–154.
31. Holmes, A. R., Cannon, R. D. & Shepherd, M. G. (1992) *Oral Microbiol. Immunol.* **7**, 32–37.
32. Smith, T. F. & Waterman, M. S. (1981) *Adv. Appl. Math.* **2**, 482–489.
33. Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995) *Cell* **80**, 631–638.
34. McDevitt, D., Francois, P., Baudaux, P. & Foster, T. J. (1994) *Mol. Microbiol.* **11**, 237–248.
35. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–147.
36. Bairoch, A. (1989) *PROSITE: A Dictionary of Protein Sites and Patterns* (Univ. of Geneva, Geneva).
37. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
38. Strynadka, N. C. & James, M. N. (1989) *Annu. Rev. Biochem.* **58**, 951–958.
39. Hostetter, M. K., Tao, N.-J., Gale, C., Herman, D. J., McClellan, M., Sharp, R. L. & Kendrick, K. E. (1995) *Biochem. Mol. Med.* **55**, 122–130.
40. Relman, D., Tuomanen, E., Falkow, S. L., Golenbock, D., Saukkonin, K. I. & Wright, S. (1990) *Cell* **61**, 1375–1382.
41. Barki, M. J., Koltin, Y., Yanko, M., Tamarkin, A. & Rosenberg, M. (1993) *J. Bacteriol.* **175**, 5683–5689.
42. Adams, A. E. M. & Pringle, J. R. (1984) *J. Cell Biol.* **98**, 934–945.
43. Hartwell, L. H. (1971) *Exp. Cell Res.* **69**, 265–276.
44. Malathi, K., Ganesan, K. & Datta, A. (1994) *J. Biol. Chem.* **269**, 22945–22951.
45. Liu, H., Kohler, J. & Fink, G. R. (1994) *Science* **266**, 1723–1726.
46. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. (1992) *Cell* **68**, 1077–1090.
47. Roberts, R. L. & Fink, G. R. (1994) *Genes Dev.* **8**, 2974–2985.
48. Saporito-Irwin, S. M., Birse, C. E., Sypherd, P. S. & Fonzi, W. A. (1995) *Mol. Cell. Biol.* **15**, 601–613.

Genomics 56, 169–178 (1999)
Article ID geno.1998.5707, available online at http://www.idealibrary.com on IDEAL®

Cloning and Characterization of a Novel β Integrin-Related cDNA Coding for the Protein TIED ("Ten β Integrin EGF-like Repeat Domains") That Maps to Chromosome Band 13q33: A Divergent Stand-Alone Integrin Stalk Structure

Randal W. Berg,* Euphemia Leung,* Sheryl Gough,† Christine Morris,† Wei-Ping Yao,* Shu-Xia Wang,‡ Jian Ni,‡ and Geoffrey W. Krissansen*,[1]

*Department of Molecular Medicine, Faculty of Medicine and Health Science, The University of Auckland, Auckland, New Zealand; ‡Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, Maryland 20850; and †Cytogenetics and Molecular Oncology Unit, Christchurch School of Medicine, Christchurch, New Zealand

Received September 2, 1998; accepted December 7, 1998

Herein we describe the cDNA sequence of a novel human gene, *ITGBL1*, encoding a β integrin-related protein termed TIED [for ten β integrin epidermal growth factor (EGF)-like repeat domains]. Overlapping cDNA clones from fetal lung, HUVEC, and osteoblast cDNA libraries encode a sequence comprising a typical signal peptide, followed by a hydrophilic 471-amino-acid domain containing 10 tandem EGF-like repeats strikingly similar to those found in the cysteine-rich "stalk-like" structure of integrin β subunits. The EGF-like repeats of TIED and β integrins are unique in that they alternate in homology and possess two additional cysteines (eight in total) whose positions differ from those in the other eight-cysteine EGF-like domains of laminin, fibrillin, and the latent TGF-β binding proteins. TIED mRNA transcripts of 2.8 kb were detected in aorta, thymus, and osteogenic sarcoma cells. The *ITGBL1* gene was mapped to human chromosome 13, band 13q33. We suggest that *ITGBL1* may be linked in some way with the evolution of the integrin β subunits. © 1999 Academic Press

INTRODUCTION

Integrins are a superfamily of dimeric αβ cell-surface glycoproteins that mediate the adhesive functions of many cell types, enabling cells to interact with one another and with the extracellular matrix (ECM) (reviewed by Hynes, 1992). Electron microscopy reveals that integrins have a globular ligand-binding head composed of parts of both subunits and two stalks that extend to the plasma membrane (Carrell *et al.*, 1985; Nermut *et al.*, 1988). All eight identified integrin β subunits are highly similar (31–46% amino acid identity), where the stalk region is composed of a fourfold repeat of a cysteine-rich segment that is thought to be internally disulfide-bonded. No function has been ascribed to the stalk region, apart from the fact that it probably facilitates ligand binding by ensuring that the globular head extends beyond the glycocalyx. The stalk region appears to be a conduit for signaling events that either lead to integrin activation or are induced in response to ligand binding. Thus the AG89 mAb preferentially recognizes the cysteine repeat region following integrin activation and can itself induce activation of β1-integrin (Takagi *et al.*, 1997).

A previous comparison had revealed that the integrin β subunit cysteine-rich repeats were homologous with a cysteine-rich repeat region in domain III of laminin B chains (Yuan *et al.*, 1990). The four cysteine-rich repeats in β integrin subunits were most related to the first four repeats in domain III (20–40%). Part of the repeat unit of the laminin B1 chain was shown to contain a sequence similar to an EGF domain; however, the cysteine repeats in laminin are larger than those of EGF and contain eight rather than six cysteine residues (Pikkarainen *et al.*, 1988). Pairwise sequence identity comparisons between EGF modules of different proteins suggest that the laminin EGF repeats, and hence also the integrin repeats, are "outliers" and should be described as EGF-like until 3D structural comparisons can confirm their family membership (Campbell and Bork, 1993).

EGF-like domains contained in many growth factors, receptors, adhesion molecules, and proteins of the coagulation and fibrinolytic pathway have either been shown or are expected to participate in protein–protein or protein–cell interactions (Campbell and Bork, 1993; Appella *et al.*, 1988; Engels 1989). Interestingly, EGF domains in several proteins, including the integrin β5

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF072752.

[1] To whom correspondence should be addressed. Telephone: (64) 9 373-7599, Ext 6280. Fax: (64) 9 373 7492. E-mail: gw.krissansen@auckland.ac.nz.

0888-7543/99 $30.00

(Ramaswamy and Hemler, 1990) and β6 (Sheppard *et al.*, 1990) subunits, the laminin-associated protein nidogen (Timpl *et al.*, 1990), the glycoproteins PAS-6/7 (Andersen *et al.*, 1997) and lactadherin (Taylor *et al.*, 1997), and entactin (Dong *et al.*, 1995), contain the small tripeptide RGD, which is a major integrin binding site. Thus PAS-6/7 and lactadherin bind the integrin αvβ5 in an RGD-dependent fashion, and the RGD motifs in entactin bind αvβ3 and possibly α3β1. These EGF domains may participate in integrin-mediated RGD-dependent cell adhesion events. The site in laminin that mediates cell attachment, migration, and receptor binding was localized to the peptide CDPGYIGSR in the EGF-like repeat domain III of the B1 chain (Graf *et al.*, 1987). EGF domains in some ECM proteins are mitogenic as exemplified by those in the inner short arm structures of laminin (Panayotou *et al.*, 1989).

Here we report the cDNA sequence of a new member of the EGF-like protein family, termed TIED, that has the potential to provide novel insights into the evolution and alternative functions of the stalk structure of integrin β subunits.

MATERIALS AND METHODS

Cell culture. The human osteogenic sarcoma U-2OS cell line (Ponten and Saksela, 1967) obtained from the American Type Culture Collection (ATCC) was cultured in McCoy's 5A medium supplemented with 10% FBS, 2 mM glutamine, 50 μg/ml penicillin, and 50 μg/ml streptomycin, at 37°C in a 5% CO_2 atmosphere.

Screening of cDNA libraries. The TIED cDNA was initially identified as an expressed sequence tag (EST) following screens for integrin homology in an EST cDNA database using the BLAST network service provided by the National Center for Biotechnology Information. Partial-length TIED cDNA clones HSRAZ62 and HLHFV34 were identified in databases from human osteoclastoma and fetal lung cDNA libraries, respectively. Further clones were identified by screening fetal lung and umbilical vein endothelial cDNA libraries constructed using the LambdaZAP II vector (Stratagene, La Jolla, CA). Libraries were replica plated onto Gene Screen Plus filters (DuPont, Boston, MA), and screened as described previously (Yuan *et al.*, 1992) using either a 900-bp *Eco*RI/*Eco*RI fragment from clone HLHFV34 or a 223-bp PCR product, encompassing nucleotides 1216 to 1438, generated by PCR with the primers 62F 5′-ATGACGGAAGAACAAAGCAAGAA-3′ and 62R 5′-ATCCATCCCAGCAATCACAGTT-3′ from clone HSRAZ62.

DNA sequencing. DNA sequences were determined by cycle sequencing using an Applied Biosystems 373A automated DNA sequenator (The Centre for Gene Technology, School of Biological Sciences, University of Auckland, Auckland, New Zealand). The composite TIED sequence was obtained on both strands of the overlapping cDNA clones HSRAZ62, HLHFV34, S0003.9, HOHCH55, and HUVEC5.1.1, using a combination of Universal M13 and sequence-specific primers. Sequence analysis was performed using the Wisconsin package version 9.1 from the Genetics Computer Group (GCG) (Madison, WI).

Polymerase chain reaction (PCR). The expression of *ITGBL1* was analyzed by PCR using DNA templates from a human thymus cDNA library (ATCC) and cDNA prepared from mRNA extracted from U-2OS cells. Thermocycling parameters were 94°C for 1 min; 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s; followed by a final extension at 72°C for 3 min. For the chromosomal assignment, PCR was initially carried out with 24 cell hybrid DNAs in which 14 of the hybrids contained a single chromosome and the remaining 10 contained two to three chromosomes or one to three chromosomal fragments (Kelsell *et al.*, 1995). Subsequently a set of 7 cell hybrid DNAs was employed, in which cell hybrids MOG34A4, DUR4.3, SIR74ii, and LSR34S49 contain chromosome 13, and hybrids TWIN19-D12, CTP34B4, and DT1.2.4 together contain all other human chromosomes except for chromosomes 13, 9, and 19. Two primers, GM-F (5′-CGAATGAAATCCGAGTACCTATTAG-3′) and GM-R (5′-GCATCCCTGGCCTCTACCCAC-3′), were designed to amplify a region encompassing nucleotides 1618 to 1839 of the TIED cDNA sequence. They amplified a PCR product of 222 bp from human DNA, but not from mouse or hamster DNA. The PCR conditions for amplification from cell hybrid DNAs were as above except that annealing was carried out at 62°C and extension lasted for 45 s. PCR products were resolved on 2% agarose gels, stained with ethidium bromide, and transferred to GeneScreen Plus. Blots were hybridized with a ^{32}P-labeled 2.5 kb *Not*I/*Not*I fragment of clone HOHCH55 in 5× SSC, 5× Denhardt's solution, 50% formamide, with 1% SDS and 100 μg/ml denatured salmon sperm DNA, at 42°C. They were washed twice in 0.1× SSC, 0.1% SDS at 60°C for 30 min and autoradiographed.

Fluorescence in situ hybridization (FISH). Metaphase spreads were prepared from phytohemagluttinin-stimulated peripheral blood lymphocytes of a 46,XY male donor using standard cytogenetic procedures. The 2.5-kb insert of clone HOHCH55 was labeled with biotin-16-dUTP using a Biotin High Prime labelling kit (Boehringer Mannheim). Conditions for hybridization and immunofluorescent detection were essentially as described (Morris *et al.*, 1993), except that Cot 1 suppression was not required, slides were washed with 0.1× SSC at 60°C, and an additional amplification step was included. For precise chromosome band localization, DAPI and FITC images were captured separately for each metaphase from the fluorescence microscope using a Photometrics KAF1400 CCD camera and QUIPS

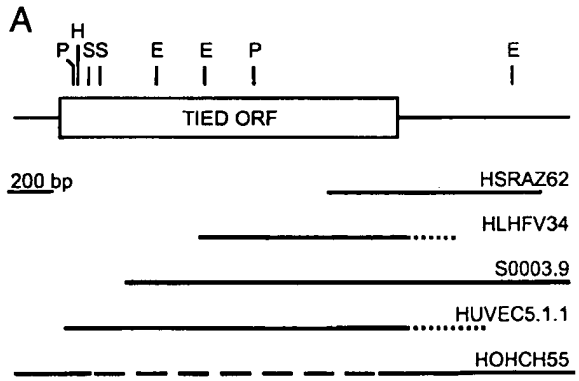


FIG. 1. Alignment of TIED cDNA clones, and their nucleotide and deduced amino acid sequences. (**A**) Partial restriction map of the composite TIED cDNA sequence and alignment of cDNA clones. The schematic at the top shows the open reading frame (ORF) as an open box, and 5′- and 3′-untranslated regions are shown as solid lines. The positions of recognition sites for the restriction enzymes *Pvu*II (P), *Hin*dIII (H), *Sma*I (S), and *Eco*RI (E) are indicated by vertical lines. The relative positioning of the TIED cDNA clones, which were sequenced on both strands, is indicated at the bottom with solid lines. The dashed region in HOHCH55 was sequenced on one strand only. The dotted 3′-untranslated regions in HLHFV34 and HUVEC5.1.1 denote sequence variation with corresponding regions in HSRAZ62, S0003.9, and HOHCH55. The scale bar indicates 200 bp. (**B**) Nucleotide and deduced amino acid sequences of the TIED molecule. The numbers in the left margin refer to nucleotide and amino acid positions. The first nucleotide of the start codon and the initiator methionine have each been assigned position 1. The 10 cysteine-rich integrin β- and EGF-like repeats are indicated with solid lines below the aa sequence, and the Cys residues in each repeat are numbered after Yuan *et al.* (1990). The stop codon is represented by an asterisk, the putative signal peptide and polyadenylation signal site (AATAAA) are underlined, and a potential site for N-linked glycosylation at position 405 is indicated (⊗).

B

```
-219                                                                ACCAGCACCCCGCCCAGAGCAGTGCCGCTGCCCAAATCC
-180 TCGCAGGCAGCTCATCAACGCAATTGCAACTCCGGCTGGAGCCCCGGACCTGCAAGCCTGGGTGTCCGTGGGTCCGTCTGCCCAGCCATC
-90  TGCTGGTGGCACCTCTCCCTCCTGCCGCCTCCCTCGGTGAACCCCACCTTGCAGAAGTGCAGCTCGCCCGGAGCAGCCCAGGAGCTCAGC
1    ATGCGTCCCCCAGGCTTCAGGAACTTCTTGCTGCTGGCGTCCTCCCTTCTCTTTGCTGGGTTGTCAGCTGTTCCTCAAAGCTTCTCGCCA
1    M  R  P  P  G  F  R  N  F  L  L  L  A  S  S  L  L  F  A  G  L  S  A  V  P  Q  S  F  S  P

91   TCTCTGAGGAGCTGGCCGGGCGCCGCCTGCAGGCTGTCCCGGGCCGAGTCCGAGCGACGCTGCCGCGCACCTGGGCAGCCCCCGGGGGCC
31   S  L  R  S  W  P  G  A  A  C  R  L  S  R  A  E  S  E  R  R  C  R  A  P  G  Q  P  P  G  A
                                                                    1———————————————————————
181  GCGCTGTGCCACGGCCGGGGCCGCTGCGACTGCGGCGTCTGCATCTGCCACGTGACTGAGCCGGGCATGTTCTTCGGGCCCCTGTGTGAG
61   A  L  C  H  G  R  G  R  C  D  C  G  V  C  I  C  H  V  T  E  P  G  M  F  F  G  P  L  C  E
     ———2———————————3———4———5———6———————————————————————————————————————————7———
271  TGCCATGAGTGGGTGTGCGAGACCTACGACGGGAGCACCTGTGCAGGCCATGGTAAGTGTGACTGTGGCAAGTGCAAGTGTGACCAGGGA
91   C  H  E  W  V  C  E  T  Y  D  G  S  T  C  A  G  H  G  K  C  D  C  G  K  C  K  C  D  Q  G
     8         1———————————————2———————————3———4———————5———6———————
361  TGGTATGGGGATGCTTGCCAGTACCCAACTAACTGTGACTTGACAAAGAAGAAAAGTAACCAAATGTGCAAGAATTCACAAGACATCATC
121  W  Y  G  D  A  C  Q  Y  P  T  N  C  D  L  T  K  K  K  S  N  Q  M  C  K  N  S  Q  D  I  I
     ———————7———————8                                                  1———————————————
451  TGCTCTAATGCAGGTACATGTCACTGTGGCAGGTGTAAGTGTGATAATTCAGATGGAAGTGGACTTGTGTATGGTAAATTTTGTGAGTGT
151  C  S  N  A  G  T  C  H  C  G  R  C  K  C  D  N  S  D  G  S  G  L  V  Y  G  K  F  C  E  C
     2———————————3———4———5———6———————————————————————————————————————————7———8
541  GACGATAGAGAATGCATAGACGATGAAACAGAAGAAATATGTGGAGGCCATGGGAAGTGTTACTGTGGAAACTGCTACTGCAAGGCTGGT
181  D  D  R  E  C  I  D  D  E  T  E  E  I  C  G  G  H  G  K  C  Y  C  G  N  C  Y  C  K  A  G
               1———————————————2———————————3———4———————5———6———————
631  TGGCATGGAGATAAATGTGAATTCCAGTGCGATATCACCCCCTGGGAAAGCAAGCGAAGATGCACGTCTCCAGATGGCAAAATCTGCAGT
211  W  H  G  D  K  C  E  F  Q  C  D  I  T  P  W  E  S  K  R  R  C  T  S  P  D  G  K  I  C  S
     ———————7———————8                                                  1———————————————2———
721  AACAGAGGGACTTGTGTATGTGGTGAATGTACCTGTCACGATGTTGATCCGACTGGGGACTGGGGAGATATTCATGGGGACACCTGTGAA
241  N  R  G  T  C  V  C  G  E  C  T  C  H  D  V  D  P  T  G  D  W  G  D  I  H  G  D  T  C  E
     ———————3———4———5———6———————————————————————————————————————————7———
811  TGTGATGAGAGGGACTGTAGAGCTGTCTATGACCGATATTCTGATGACTTCTGTTCAGGTCATGGACAGTGTAATTGCGGAAGATGTGAC
271  C  D  E  R  D  C  R  A  V  Y  D  R  Y  S  D  D  F  C  S  G  H  G  Q  C  N  C  G  R  C  D
     8         1———————————————2———————————3———4———5———
901  TGCAAAGCAGGCTGGTATGGGAAGAAGTGTGAGCACCCACAGTCCTGCACGCTGTCAGCTGAGGAGAGCATCAGGAAGTGCCAGGGAAGC
301  C  K  A  G  W  Y  G  K  K  C  E  H  P  Q  S  C  T  L  S  A  E  E  S  I  R  K  C  Q  G  S
     6———————————————7———————8                                                  1———————
991  TCGGATCTGCCTTGCTCTGGGAGGGGTAAATGTGAATGTGGCAAATGCACCTGCTATCCTCCAGGAGATCGCCGGGTGTATGGCAAGACT
331  S  D  L  P  C  S  G  R  G  K  C  E  C  G  K  C  T  C  Y  P  P  G  D  R  R  V  Y  G  K  T
     ———————2———————3———4———5———6———————————————————————————————————————————
1081 TGTGAGTGTGATGATCGCCGCTGTGAAGACCTCGATGGTGTGGTCTGTGGAGGCCACGGCACATGTTCCTGTGGTCGCTGTGTTTGTGAG
361  C  E  C  D  D  R  R  C  E  D  L  D  G  V  V  C  G  G  H  G  T  C  S  C  G  R  C  V  C  E
     7———8         1———————————————2———————————3———4———————4———6———
1171 AGAGGATGGTTTGGAAAGCTCTGCCAACATCCGCGGAAGTGTAACATGACGGAAGAACAAAGCAAGAATCTGTGTGAATCAGCAGATGGC
391  R  G  W  F  G  K  L  C  Q  H  P  R  K  C  N  M  T  E  E  Q  S  K  N  L  C  E  S  A  D  G
     ———————————————7———————8 ⊗                                        1———————————————
1261 ATATTGTGCTCGGGGAAGGGTTCTTGTCATTGTGGGAAGTGCATTTGTTCTGCTGAAGAGTGGTATATTTCTGGGGAGTTCTGTGACTGT
421  I  L  C  S  G  K  G  S  C  H  C  G  K  C  I  C  S  A  E  E  W  Y  I  S  G  E  F  C  D  C
     ———2———————————3———4———————5———6———————————————————————————————————————7———8
1351 GATGACAGAGACTGCGACAAACATGATGGTCTCATTTGTACAGGGAATGGAATATGTAGCTGTGGAAACTGTGAATGCTGGGATGGATGG
451  D  D  R  D  C  D  K  H  D  G  L  I  C  T  G  N  G  I  C  S  C  G  N  C  E  C  W  D  G  W
               1———————————————2———————————3———4———————5———6———————
1441 AATGGAAATGCATGTGAAATCTGGCTTGGCTCAGAATATCCTTAACAATTACATGAGAGAGGTCTGGATTCTTATTTTTTCTGGGCCATT
481  N  G  N  A  C  E  I  W  L  G  S  E  Y  P  *
     ———————————7
1531 AGAACATATAAATGCGAAGGAAACCATGTATATTCACCACTAGGACAGGTTAAAAAGACCATTGTATGTTTTTCTATTTCTGAATTACGA
1621 ATGAAATCCGAGTACCTATTAGAAATGAGTTATGCAAATTTAGATGCAAATAACATTAGAAAAAAAAGATTCTTCCATAATTAACATAAG
1711 TGGTTCCTAACGAGAGCAATTTTTCCACCCAAAAGTCATTTGGCAACATCTACAGACAATTTTGATTGTCACACTGGGTCGGGTAGGAAG
1801 GTATGCTGCAGACATTTGGTGGGTAGAGGCCAGGGATGCTGCTGAGCATCCCGCAGTGTACAGGACAGCCCCCAAACAAGGAATTATCCA
1891 GCCCCAAATGCCAATAGGGCTCAAACTGAGAAACATTGAGTTATATGGCTATTAGAAATCCACATTCTTACACAAGAAAGACCATATTAG
1981 AATCTAAGGAAAACATGCATATTCACATTAATTAATCGATCAGATTTTTCCAGAATTCCGTATCAGTCACCATTTTAATATGGGGACAAT
2071 GAAGACAAGCACACAGGAGGTAGAATATCAGAGTGGGGCTGGATCAAGGGCAAAAACTGGTCATTAAGTCATCTGACATTAAATCATTTA
2161 GCCACTAAGTTATTTGTGTACTCTCACTTTAAACTCACCAAAGAAGATTCTCTTAAAGAAATTATGAAAAATGTACAATTTAACATTTTA
2251 AATAAATAGTGACAGAAGTTGTTTAAAAA
```

FIG. 1—*Continued*

(Vysis Inc., Downers Grove, IL) Smartcapture FISH software (version 1.3). QUIPS CGH/Karyotyping software (version 3.0.2) assisted in karyotype analysis.

Northern blot analysis. Human MTN I and II filters and a Human RNA Master Blot (Clontech) were screened with the ^{32}P-labeled 900-bp *Eco*RI/*Eco*RI fragment of the insert of clone HLHFV34. Hybridization was carried out at 60°C in ExpressHyb solution (Clontech). Filters were washed twice in 0.1× SSC, 1% SDS at 50°C for 30 min and autoradiographed. Total RNA was isolated from the osteogenic sarcoma cell line U-2OS as described (Chomczynski and Sacchi, 1987), separated on 1% agarose formaldehyde gels, and transferred to GeneScreen Plus. Blots were hybridized with the ^{32}P-labeled 2.5-kb insert of clone HOHCH55 in 5× SSC, 5× Denhardt's solution, 50% formamide, with 1% SDS and 100 μg/ml denatured salmon sperm DNA, at 42°C. They were washed twice in 0.1× SSC, 0.1% SDS at 50°C for 30 min and autoradiographed.

RESULTS AND DISCUSSION

Cloning of a Novel Integrin β Subunit-Related Molecule

A homology search (Altschul *et al.*, 1990) of a human EST cDNA database generated through the combined efforts of Human Genome Sciences, Inc. and The Institute for Genomic Research (Adams *et al.*, 1995; Feng *et al.*, 1996), using the known amino acid sequences of

integrin subunits, identified clones HSRAZ62 and HL-HFV34 from osteoclastoma and fetal lung cDNA libraries, respectively, which represented a potential novel integrin β subunit. The HSRAZ62 clone was sequenced on both strands, and alignment of the translated sequence with integrin β subunit sequences revealed that it encoded two complete cysteine repeat domains highly similar to those contained in the β integrin stalk-like structure. However, no N-terminal methionine initiation codon was present, and the last cysteine repeat was not followed by a transmembrane domain, as in integrin β subunits. To isolate the full-length sequence for this unusual clone, a 223-bp HSRAZ62-derived PCR product (refer to Materials and Methods) was used to screen a variety of cDNA libraries including two prepared from human fetal lung and umbilical vein endothelial cells, from which positive clones were obtained. Clone S0003.9 from the fetal lung library and clone HUVEC5.1.1 from the endothelial cell library both extended the HSRAZ62 sequence, and a subsequent screen of the EST database identified the potential full-length cDNA clone HOHCH55 from an osteoblast cell cDNA library (Fig.1A).

Structure of the TIED ("Ten β Integrin EGF-like Repeat Domains") Molecule

The nucleotide and deduced amino acid sequence of the complete TIED molecule derived from the composite cDNA is shown in Fig. 1B. The 2493-nucleotide sequence includes 219 nucleotides of 5′-untranslated sequence, a 1485-nucleotide open reading frame encoding 494 amino acid residues, and 789 nucleotides of 3′-untranslated sequence that includes a consensus AATAAA poly(A) signal followed 18 nucleotides later by a poly(A) stretch. The presumptive methionine translational start codon is flanked by sequence that resembles but is not identical to a classical Kozak consensus, PurNNAUGPur. Nevertheless it is followed by a hydrophobic stretch of 23 amino acid residues that is typical of a signal peptide sequence (Fig. 2A). A recently submitted EST from the Washington University–NCI Human EST Project (Accession No. AA417383) extends the HOHCH55 sequence by 59 nucleotides and incorporates an in-frame stop codon, rendering it unlikely that the open reading frame extends upstream of the designated start codon. The putative signal peptide is followed by a predominantly hydrophilic domain of 471 amino acid residues, containing 10 EGF-like cysteine-rich repeats. The last repeat is incomplete, missing the C-terminal cysteine. The predicted molecular mass of an unglycosylated form of the mature protein is 51.4 kDa; however, there is one potential N-linked glycosylation site, Asn 405.

As this work was nearing completion, a BLAST search of the GenBank database revealed an entry, AB008375, whose sequence was essentially identical to that of TIED, except that it contains an extra G residue at nucleotide position 337, which alters the reading

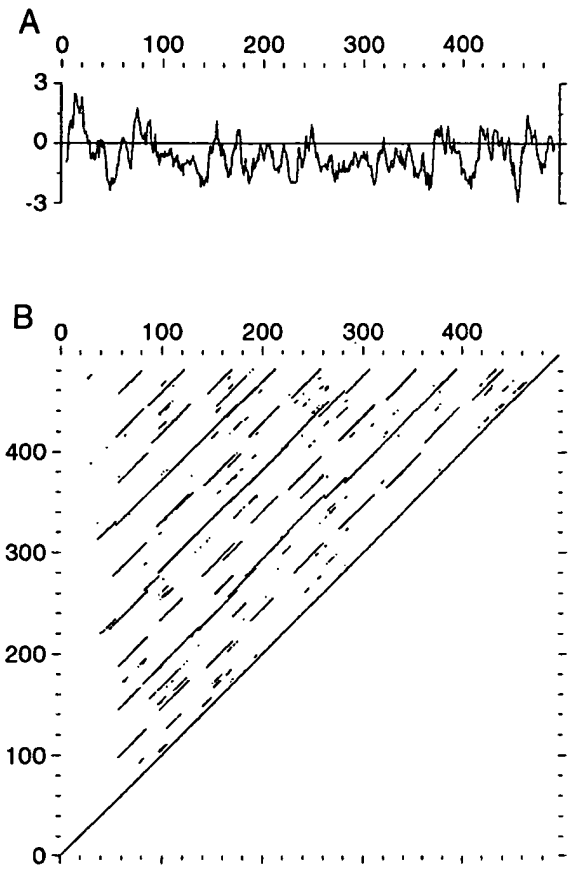


FIG. 2. Hydropathy and internal similarity plots of the TIED molecule. **(A)** Hydropathy plot (Kyte and Doolittle, 1982) of the deduced TIED protein, illustrating the hydrophobicity of the first 23 amino acid residues predicted to represent a functional signal peptide. **(B)** Dot-matrix comparison illustrating the repetitive nature of the deduced amino acid sequence of TIED. In this GCG plot, amino acid residues in the TIED sequence are compared with one another in pairwise fashion. Similarities are converted to dots that form clusters and diagonal lines, with complete identity along the central diagonal.

frame. Thus the encoded molecule is N-terminally truncated, being identical to TIED C-terminal to amino acid residue Gly 113, but extending only 21 residues further N-terminal. In addition to several nucleotide substitutions, AB008375 harbors a 68-bp deletion (nucleotides 145 to 212). The encoded molecule was proposed to be osteoblast-specific, but this seems unlikely given our expression data for TIED.

The TIED EGF-like Domains Are Remarkably Similar to Those of Integrin β Subunits

The repetitious modular structure of TIED is most clearly illustrated in Fig. 2B by a dot-matrix comparison, where the presence of repeats is visualized by lines and dashes that run parallel to the central diagonal that marks amino acid identity. Comparison of the deduced TIED sequence with EGF-like proteins in the GenBank database revealed that the TIED repeats were most similar with the β integrin cysteine-rich repeats. There are two features that distinguish the integrin β subunit and TIED repeats from the majority of other EGF-like proteins. The EGF domains of TIED

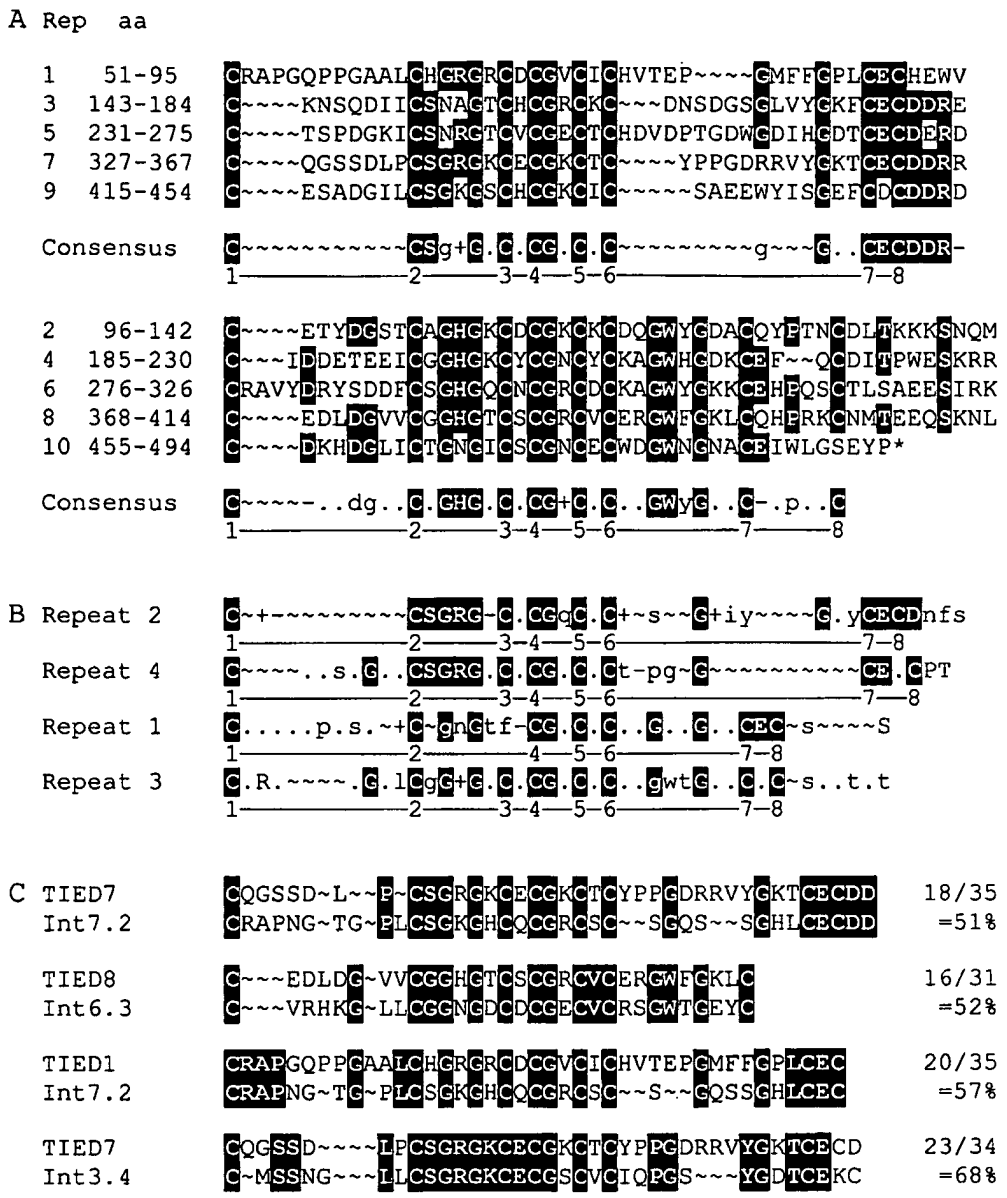


FIG. 3. The EGF-like domains of TIED are highly similar to those of β integrins. **(A)** The EGF-like domains of TIED alternate in homology. Amino acid sequences of the odd (upper) and even (lower) numbered EGF-like repeats in TIED are aligned, with consensus sequences shown below each group. Consensus amino acid residues are boxed in black. Numeration of amino acid residues is as in Fig. 1B. **(B)** The EGF-like domains of the β integrins also alternate in homology. Consensus sequences of the four EGF-like repeats of the human integrin β1 through β8 subunits have been aligned. Identical amino acid residues are boxed in black. **(C)** Similarity between the TIED and the β integrin EGF-like domains extends to large stretches of sequence identity. The first, seventh, and eighth EGF-like repeats of TIED have been aligned and compared with the second repeat of integrin β7 (Int7.2), the fourth repeat of integrin β3 (Int3.4), and the third repeat of integrin β6 (Int6.3), respectively. Identical amino acid residues are boxed in black. The percentage identity is indicated in the right margin. Gaps introduced to optimize the alignments are denoted by tildes (~). In the consensus sequences, uppercase letters indicate that 75 to 100% of the sequences aligned contain the amino acid at that position; lowercase letters indicate 50 to 75%; plus and minus signs indicate basic and acidic residues, respectively; and each dot represents the position of an amino acid residue that is not conserved. Conserved Cys residues are numbered as in Fig. 1B or after Yuan *et al.* (1990).

and β integrins contain eight cysteines, rather than the six cysteines found in the "classical" EGF domain (Figs. 3A and 3B). Only the first integrin repeat and the last TIED repeat are exceptions to the rule, possessing seven cysteines, as found in some of the EGF-like modules of fibrillin that have been termed "hybrid" domains (Corson *et al.*, 1993; Pereira *et al.*, 1993). Other members of the EGF-like family that possess eight-cysteine repeats include laminin (Pikkarainen *et al.*, 1988), fibrillin (Corson *et al.*, 1993; Pereira *et al.*, 1993), and the latent TGF-β binding proteins (LTBP) (Kanzaki *et al.*, 1990; Morén *et al.*, 1994; Tsuji *et al.*, 1990; Yin *et al.*, 1995). The positions of the "extra" two cysteine residues are unique to TIED and the β integrins and are not conserved in other eight-cysteine EGF-like domains present in fibrillin, laminin, and LTBPs (Fig. 4).

Second, the repeats show alternating similarity, such that odd-numbered repeats are most similar to one another, and vice versa, the even-numbered repeats are more similar to one another than they are to the odd-numbered repeats. To our knowledge this par-

```
EGF       CPLSHDGYCLHDGVCMYI~~EALDKYACNCVV~~~~~~GYIGERCQ

LamB1     CECHLQ~~GSVMAFCNPV~~~~~~TGQCHCFQ~~~~~~GVYARQCDRCLP~~GWWGFPSCQP

LTBP-2      CYR~~~SLGPGTCTLPLAQRITKQICCCSRVG~~~~KAWGSECEKCPLPGTEAFREICC
Fibn        CFL~~~RYEDEECTLPIAGRHRMDACCCSVGA~~~~AWGTEECEECPMRNTPEYEELCC

TIED9       CES~~~~ADGILCS~~GKGSCHCGKCICS~~~AEEWYISGEFCDCDDRD
Int2.2      CRK~~~DNNSIICS~~GLGDCVCGQCLCHTSDVPGKLIYGQYCECDTIN
TIED6       CRAVYDRTSDDFCS~~GHGQCNCGRCDCK~~~~~~AGWYGKKCEHPQSCTLSAEESIRD
Int2.3      CER~~~~YNGQVCGGPGRGLCFCGKCRCH~~~~~~PGFEGSACQCERTTEG
```

FIG. 4. Comparison of the cysteine "footprint" of the TIED and β integrin EGF-like repeats with those in other proteins harboring eight cysteine EGF-like domains. The eight cysteines contained in representative EGF-like repeats found in the laminin B1 chain (LamB1), latent TGF-β binding protein (LTBP-2), and fibrillin (Fibn) have been aligned with cysteines in the sixth and ninth repeats of TIED and the second and third repeats of the integrin β2 subunit (Int2.2 and Int2.3, respectively). The eight cysteine-repeat motifs have been compared with the six-cysteine EGF repeat motif. Gaps introduced to optimize the alignments are denoted by tildes (~). Cysteine residues are boxed in black.

ticular feature is shared only by the β integrins and the TIED molecule. A schematic comparison of the TIED and β integrin structures is shown in Fig. 5. The odd-numbered TIED repeats are most similar to the even-numbered β integrin repeats. In particular the sequence CSGRG is highly conserved, as is the CECD sequence (except for the fourth integrin repeat) (Figs. 3A and 3B). The number of amino acids intervening between cysteines at positions 6 and 7 in these repeats varies markedly for both molecules. Vice versa, the even-numbered TIED repeats are most similar to the odd-numbered β integrin repeats, although this is not quite as obvious since the odd-numbered β integrin repeats appear to have diverged significantly during evolution. Importantly, the similarity between the β integrin and the TIED repeats does not relate just to conserved cysteine and glycine residues, but in some regions extends across the entire sequences. Comparison of the second repeat of the integrin β7 subunit with the first and seventh TIED repeats reveals 57 and 51% identity over 35 amino acid residues, respectively, which increases to 66% similarity when conservative substitutions are taken into account (Fig. 3C). Likewise, the fourth repeat of the integrin β3 subunit shares 68% amino acid identity over 34 amino acid residues with the seventh TIED repeat, and the third repeat of the integrin β6 subunit shares 52% amino acid identity with the eighth TIED repeat over 31 amino acid residues. Interestingly, EGF domains in several proteins contain the small tripeptide RGD, which is a major integrin binding site. The TIED sequence does not include an RGD motif or other common integrin binding motifs.

A class of EGF repeats found in functionally diverse proteins contain Ca^{2+} binding domains that have the consensus sequence Asp/Asn-x-Asp/Asn-Glu/Gln-x_n-Asp/Asn*-x_n-Phe/Tyr (where n is variable, and the asterisk indicates possible β-hydroxylation). Solution structures suggest that a conserved aromatic residue in a Gly-Aromatic-x-Gly motif between Cys 5 and 6 (Downing *et al.*, 1996; Rao *et al.*, 1995) and Ca^{2+} ions (Knott *et al.*, 1996) are both key elements involved in interdomain interactions that stabilize the three-dimensional structure of EGF modules. Some of the odd-numbered TIED repeats and the second β integrin repeat have the sequence Glu/Asp-x-Asp-Asp/Glu/Gln (where x is the eighth cysteine residue), resembling part of the core Ca^{2+} binding sequence.

Alternative Splicing of TIED 3'-Untranslated Regions

The 3'-ends of cDNA clones HUVEC5.1.1 and HLHFV34 diverge from clones HSRAZ62, S0003.9, and HOHCH55 at nucleotide positions 1476 and 1502, respectively (Fig. 6A). To determine whether the 3'-untranslated region might undergo alternative splicing, PCR primers were designed to the alternative 3'-untranslated regions and used to amplify TIED transcripts from fetal thymus cDNA. Primer pairs 62F and 57 (5'-TTTAACCTGTCCTAGTGGTG-3'; nucleotides 1565–1584) and 62F and 59 (5'-TGTCTGCAGCATACCTTCC-3'; nucleotides 1796–1814) that should amplify sequences contained in the S0003.9 and HOHCH55 cDNA clones both generated correct-sized PCR products, whereas the primer pair 62F/58 (5'-TAATGAATTCCAATGTCTGTGC-3') that should amplify

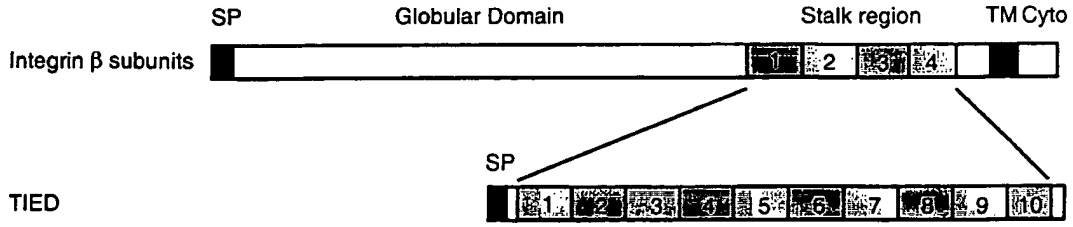


FIG. 5. Schematic comparison of the principal structural features of TIED with β integrins. The EGF-like repeats are numbered and shaded according to their alternating homology. Predominantly hydrophobic uncharged regions are denoted as solid blocks. SP, signal peptide; TM, transmembrane domain; and Cyto, cytoplasmic domain.

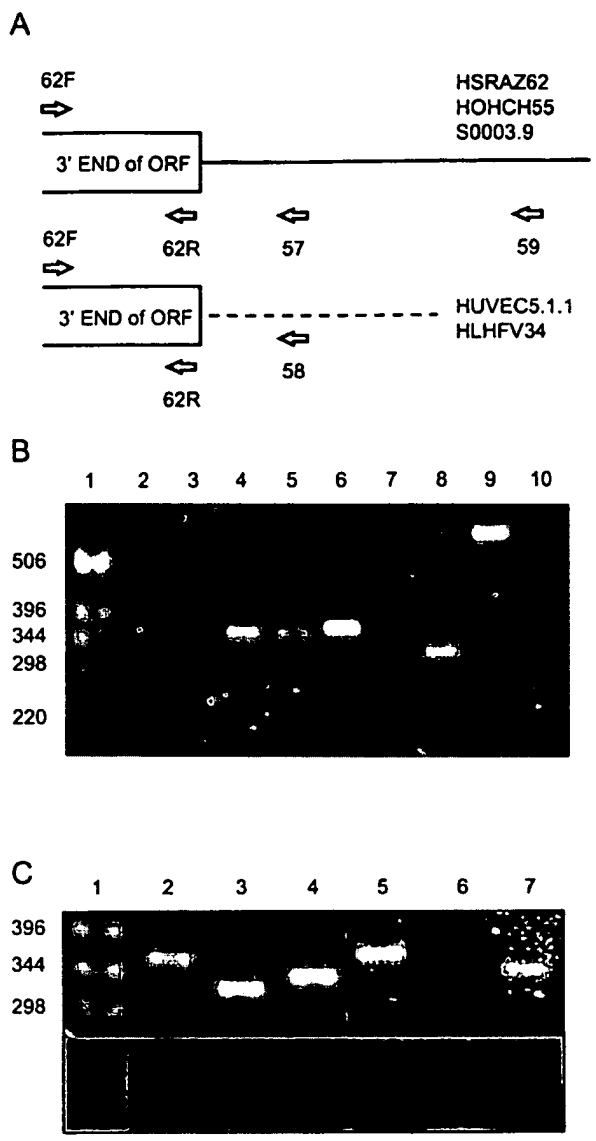


FIG. 6. TIED cDNA clones possess alternative 3′-untranslated regions: authenticating the 3′-ends by RT-PCR. To determine which of the different 3′-untranslated sequences in the various cDNA clones were authentic, RT-PCR analysis was performed using antisense primers to the alternative 3′-ends. (**A**) The locations of PCR primers are shown in the upper schematic diagram where the open reading frame (ORF) is boxed, and the 3′-untranslated region is denoted by solid or dashed lines. (**B**) PCR products obtained with primers 62F/62R (lanes **2** and **3**), 62F/57 (lanes **4** and **5**), 62F/58 (lanes **6**, **7**, and **8**), and 62F/59 (lanes **9** and **10**) were stained with ethidium bromide. Templates were plasmids containing the cDNA inserts HOHCH55 (lane **2**), S0003.9 (lanes **4** and **9**), HLHFV34 (lane **6**), and HUVEC5.1.1 (lane **8**), and cDNA from a fetal thymus library (ATCC) (lanes **3**, **5**, **7**, and **10**). A ladder of DNA markers is shown in lane **1**, with the sizes indicated in the left margin. (**C**) RT-PCR analysis of TIED transcripts in total RNA from human U-2OS osteogenic sarcoma cells (lanes **2** to **4**) and PCR from a HOHCH55 plasmid template (lanes **5** to **7**). PCR primers were 27/29 (lanes **2** and **5**), 62F/58 (lane **3** and **6**), and 62F/57 (lanes **4** and **7**). Lane **1**, DNA markers of 396, 344, and 298 bp. (**Top**) Ethidium bromide staining of PCR products; (**bottom**) the products have been hybridized to the ^{32}P-labeled insert of clone HOHCH55.

HUVEC5.1.1 and HLHFV34 sequences failed to generate a PCR product, despite producing the expected 333- and 379-bp products from the respective plasmid templates (Fig. 6B).

Since the HSRAZ62 and HOHCH55 clones were from osteoclastoma and osteoblast cDNA libraries, we examined expression of the alternative TIED transcripts in human osteogenic sarcoma U-2OS cells. RT-PCR with the primer pair 27 (5′-CTGTGGAAACTGCTACTGC-3′) and 29 (5′-CGTGCAGGACTGTGGGTGC-3′) and primer pair 62F/57 expected to amplify regions encompassing nucleotides 603 to 951 and 1216 to 1584 generated products of the expected sizes of 349 and 369 bp, respectively, which hybridized with a HO-

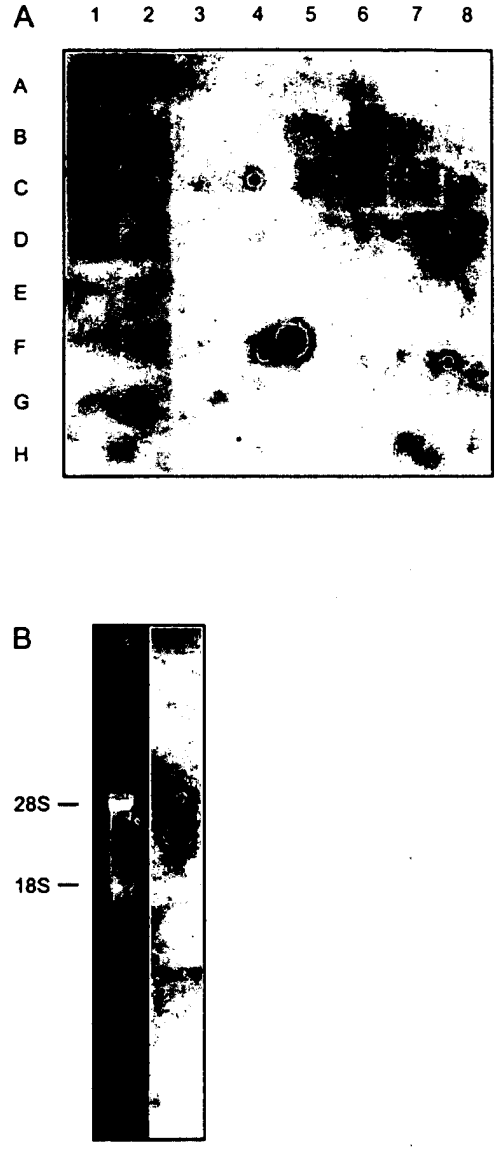


FIG. 7. Expression of TIED mRNA transcripts in various human tissues and in U-2OS osteogenic sarcoma cells. (**A**) A human RNA Master Blot (Clontech) was hybridized with a ^{32}P-labeled 900-bp *Eco*RI/*Eco*RI fragment from cDNA clone HLHFV34. The entire blot is shown illustrating detectable expression only in aorta (**C2**), whereas signals from dots containing poly(A)$^+$ RNA from 49 other tissues were not above background. The blot was rescreened to distinguish background spots from positive signals. Only the signal from aorta poly(A)$^+$ RNA was reproduced (not shown). (**B**) Northern blot of 15 μg of total RNA from U-2OS cells hybridized with the ^{32}P-labeled insert of clone HOHCH55 (right lane). The left lane illustrates an ethidium bromide stained agarose gel containing the total RNA isolated from U-2OS cells. Positions of 28S and 18S rRNAs are indicated in the left margin.

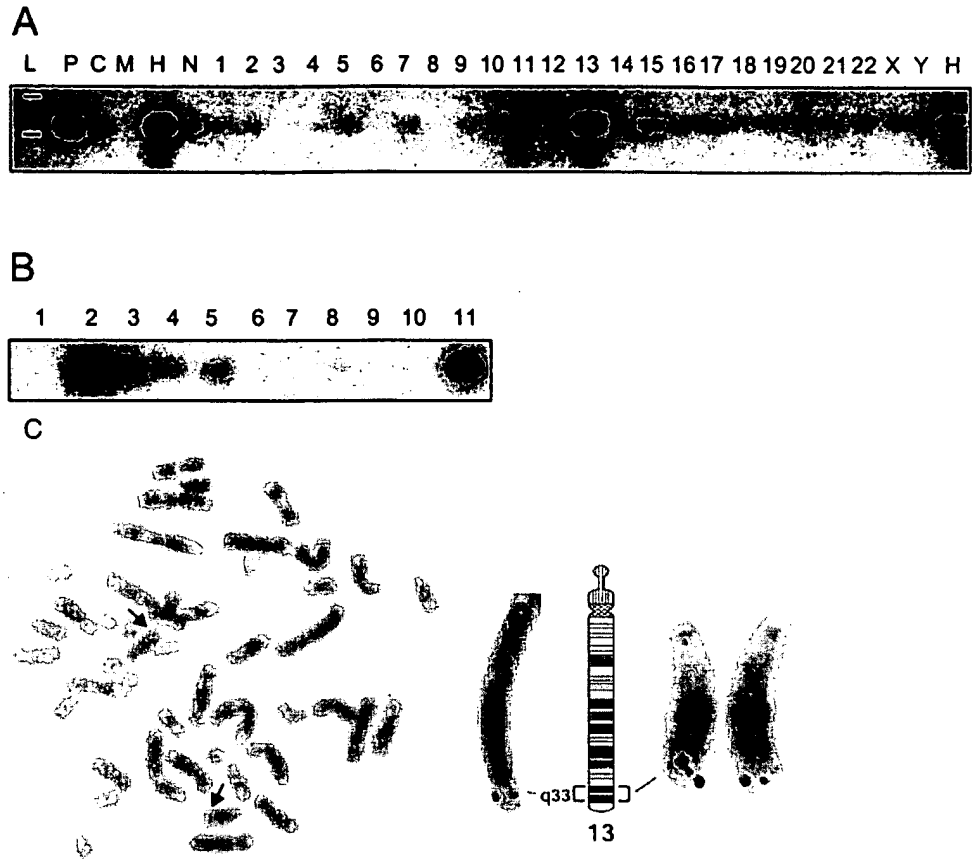


FIG. 8. The human *ITGBL1* gene maps to chromosome 13, band 13q33. (**A**) PCR analysis of a monochromosomal panel of human–rodent cell hybrid DNAs with the GM-F/GM-R primer pair generated a prominent 222-bp PCR product from chromosome 13, human DNA, and the HOHCH55 plasmid control. An autoradiograph of the PCR products hybridized with the ^{32}P-labeled insert of clone HOHCH55 is shown. Lanes correspond to DNA markers of 300 and 200 bp (**L**); plasmid HOHCH55 control (**P**); hamster (**C**), mouse (**M**), human (**H**) genomic DNA; no DNA control (**N**); and chromosome-specific somatic cell hybrid DNAs (chromosomes **1** to **22**, **X**, and **Y**). (**B**) PCR analysis of a second panel of human–rodent cell hybrid DNAs. PCR amplification with the GM-F/GM-R primer set was from DNAs of cell hybrids that contained chromosome 13 (MOG34A4, lane **2**; DUR4.3, lane **3**; SIR74ii, lane **4**; LSR34S49, lane **5**) and from hybrids that contained all other human chromosomes except for chromosomes 13, 9, and 19 (TWIN19-D12, lane **6**; CTP34B4, lane **7**; DT1.2.4, lane **8**). Control lanes include the following: lane **1**, no DNA; lane **9**, chromosome 9-specific hybrid DNA; lane **10**, chromosome 19-specific hybrid DNA; and lane **11**, chromosome 13-specific hybrid DNA. (**C**) Localization of the *ITGBL1* gene by FISH. Gray-scale inverted image of a complete metaphase cell showing fluorescent signals on chromosome 13, band q33, after hybridization of the biotinylated HOHCH55 cDNA probe (**left**). Idiogram of chromosome 13 with the q33 band bracketed and aligned with signals from enlarged copies of chromosome 13 selected from three different metaphase cells (**right**).

HCH55 cDNA probe (Fig. 6C). A 333-bp PCR product could also be amplified from U-2OS cDNA using the 62F/58 primer pair. Thus the 3′-ends of all the cDNA clones are authentic and result from alternative splicing, where those of HSRAZ62, S0003.9, and HOHCH55 are expressed in both thymus and osteogenic sarcoma cells, and those of the HUVEC5.1.1 and HLHFV34 are expressed only in the latter.

TIED mRNA Is Widely Expressed but Not Abundant

TIED cDNA clones were detected in osteoclastoma, osteoblast, umbilical vein, and fetal lung cDNA libraries, suggesting that TIED might be widely expressed; however, no clones were obtained from fetal heart and adrenal gland tumor cell derived libraries. Screening of a human RNA master blot (Clontech) containing RNAs from 50 different tissues revealed readily detectable expression of TIED mRNA transcripts only in aorta (Fig. 7A), suggesting that the TIED message is not particularly abundant in the tissues examined apart from aorta. TIED transcripts were not detected in either adult or fetal heart, suggesting that expression was specific for aorta. Northern blot analysis of total RNA prepared from U-2OS osteogenic sarcoma cells revealed a single transcript of approximately 2.8 kb (Fig. 7B).

The Human ITGBL1 Gene Maps to Chromosome 13, Band 13q33

PCR from genomic DNA of a panel of human–rodent hybrid cell lines was used to map the human *ITGBL1* gene to a particular chromosome. The expected 222-bp PCR product was specifically amplified from human genomic DNA and from the 289 hybrid, which contains human chromosome 13 and fragments of chromosome 8, 11, and 12 (Fig. 8A). *ITGBL1* sequences were not amplified from hybrids C4A, JIC14, and 1aA9602+, which contain human chromosomes 8, 11, and 12. As-

signment of the *ITGBL1* gene to chromosome 13 was confirmed by PCR analysis of a second series of chromosomal hybrids. An *ITGBL1* PCR product was amplified from the DNA of four hybrids that contained chromosome 13 (MOG34A4, DUR4.3, SIR74ii, and LSR34S49), but not from hybrid DNAs that contained all other human chromosomes except for chromosomes 13, 9, and 19 (TWIN19-D12, CTP34B4, and DT1.2.4) (Fig. 8B).

The precise localization of the *ITGBL1* gene was determined by FISH analysis using the HOHCH55 cDNA insert as a probe. Of 40 metaphase cells examined, 40 showed fluorescent signals on one or both chromosomes 13, specifically across band q33 (Fig. 8C). No additional site-specific signals were detected on any other chromosome. Other genes that have been mapped to chromosome band 13q33 include the pro alpha 1 and 2 (IV) collagen genes (Boyd *et al.*, 1988), the DNA ligase IV gene (Wei *et al.*, 1995), and the gene for xeroderma pigmentosum complementation group G (XPG) (Samec *et al.*, 1994). In terms of disease association, band 13q33 is a site for integration by human papilloma virus-33 (Gilles *et al.*, 1996); it is amplified in oral squamous cell carcinomas (Matsumura, 1995) and is commonly deleted in ovarian cancer (Yang-Feng *et al.*, 1992).

In summary, we predict that TIED is a secreted protein linked in evolution to the stalk-like structure of integrin β subunits. Whether an ancestral TIED-like molecule was integrated into β integrins via gene conversion and attributes integrins with novel functions is not known. Given that EGF-like domains participate in protein–protein and protein–cell interactions, future studies will need to appraise whether TIED protein has proadhesive, anti-adhesive, and/or growth factor activities.

ACKNOWLEDGMENTS

This work was supported in part by grants from the Wellcome Trust (UK), the Health Research Council of New Zealand, The Royal Society of New Zealand, the Lottery Grants Board of New Zealand, the Cancer Society of New Zealand Inc., the Marsden Fund, the Auckland Medical Research Foundation, the Maurice and Phyllis Paykel Trust, and the Multiple Sclerosis Society of New Zealand. C.M. is a Travis Trust Senior Cancer Research Fellow. G.W.K was a recipient of a Wellcome Senior Research Fellowship in Medical Science in New Zealand and is currently a recipient of a James Cook Research Fellowship from the Royal Society of New Zealand. The initial EST clone used in this study was discovered as part of a joint collaboration between scientists at The Institute for Genomic Research (Rockville, MD), and at Human Genome Sciences Inc. (Rockville, MD).

REFERENCES

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee N. H., *et al.* (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**(Suppl.): 3–174.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Andersen, M. H., Berglund, L., Rasmussen, J. T., and Petersen, T. E. (1997). Bovine PAS-6/7 binds alpha v beta 5 integrins and anionic phospholipids through two domains. *Biochemistry* **36:** 5441–5446.

Appella, E., Weber, I. T., and Blasi, F. (1988). Structure and function of epidermal growth factor-like regions in proteins. *FEBS Lett.* **231:** 1–4.

Boyd, C. D., Toth-Fejel, S. E., Gadi, I. K., Litt, M., Condon, M. R., Kolbe, M., Hagen, I. K., Kurkinen, M., Mackenzie, J. W., and Magenis, E. (1988). The genes coding for human pro alpha 1(IV) collagen and pro alpha 2 (IV) collagen are both located at the end of the long arm of chromosome 13. *Am. J. Hum. Genet.* **42:** 309–314.

Campbell, I. D., and Bork, P. (1993). Epidermal growth factor-like modules. *Curr. Opin. Struct. Biol.* **3:** 385–392.

Carrell, N. A., Fitzgerald, L. A., Steiner, B., Erickson, H. P., and Phillips, D. R. (1985). Structure of human platelet membrane glycoproteins IIb and IIIa as determined by electron microscopy. *J. Biol. Chem.* **260:** 1743–1749.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Corson, G. M., Chalberg, S. C., Dietz, H. C., Charbonneau, N. L., and Sakai, L. Y. (1993). Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5′ end. *Genomics* **17:** 476–484.

Dong, L-J., Hsieh, J-C., and Chung, A. E. (1995). Two distinct cell attachment sites in entactin are revealed by amino acid substitutions and deletion of the RGD sequence in the cysteine-rich epidermal growth factor repeat 2. *J. Biol. Chem.* **270:** 15838–15843.

Downing, A. K., Knott, V., Werner, J. M., Cardy, C. C., Campbell, I. D., and Handford, P. A. (1996). Solution structure of a pair of calcium-binding epidermal growth factor-like domains: Implications for the Marfan syndrome and other genetic disorders. *Cell* **85:** 597–605.

Engel, J. (1989). EGF-like domains in extracellular matrix proteins: Localized signals for growth and differentiation? *FEBS Lett.* **251:** 1–7.

Feng, G. S., Ouyang, Y. B., Hu, D. P., Shi, Z. Q., Gentz, R., and Ni, J. (1996). Grap is a novel SH3-SH2-SH3 adaptor protein that couples tyrosine kinases to the Ras pathway. *J. Biol. Chem.* **271:** 12129–12132.

Gilles, C., Piette, J., Ploton, D., Doco-Fenzy, M., and Foidart, J. M. (1996). Viral integration sites in human papilloma virus-33-immortalized cervical keratinocyte cell lines. *Cancer Genet. Cytogenet.* **90:** 63–69.

Graf, J., Iwamoto, Y., Sasaki, M., Martin, G. R., Kleinman, H. K., Robey, F. A., and Yamada, Y. (1987). Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding. *Cell* **48:** 989–996.

Hynes, R. O. (1992). Integrins: Versatility, modulation, and signaling in cell adhesion. *Cell* **69:** 11–25.

Kanzaki, T., Olofsson, A., Morén, A., Wernstedt, C., Hellman, U. K., Claesson-Welsh, L., and Heldin, C.-H. (1990). TGF-beta 1 binding protein: A component of the large latent complex of TGF-beta 1 with multiple repeat sequences. *Cell* **61:** 1051–1061.

Kelsell, D. P., Rooke, L., Warne, D., Boucyk, M., Cullin, L., Cos, S., West, L., Povey, S., and Spurr, N. K. (1995). Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. *Ann. Hum. Genet.* **59:** 233–241.

Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. (1996). Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. *J. Mol. Biol.* **255:** 22–27.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Matsumura, K. (1995). Detection of DNA amplifications and deletions in oral squamous cell carcinoma cell lines by comparative

genomic hybridization (CGH). *Kokubyo Gakkai Zasshi J. Stomatol. Soc. Jpn.* **62:** 513–531.

Morén, A., Olofsson, A., Stenman, G., Sahlin, P., Kanzaki, T., Claesson-Welsh, L., Dijke, P., Miyazono, K., and Heldin, C. H. (1994). Identification and characterization of LTBP-2, a novel latent transforming growth factor-beta-binding protein. *J. Biol. Chem.* **269:** 32469–32478.

Morris, C. M., Courtay, C., Guerts van Kassel, A., ten Hoeve, J., Heisterkamp, N., and Groffen, J. (1993). Localization of a gamma-glutamyl-transferase-related gene family on chromosome 22. *Hum. Genet.* **91:** 31–36.

Nermut, M. V., Green, N. M., Eason, P., Yamada, S. S., and Yamada, K. M. (1988). Electron microscopy and structural model of human fibronectin receptor. *EMBO J.* **7:** 4093–4099.

Panayotou, G., End, P., Aumailley, M., Timpl, R., and Engel, J. (1989). Domains of laminin with growth-factor activity. *Cell* **56:** 93–101.

Pereira, L., D'Alessio, M., Ramirez, F., Lynch, J. R., Sykes, B., Pangilinan, T., and Bonadio, J. (1993). Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. *Hum. Mol. Genet.* **2:** 961–968.

Pikkarainen, T., Kallunki, T., and Tryggvason, K. (1988). Human laminin B2 chain. Comparison of the complete amino acid sequence with the B1 chain reveals variability in sequence homology between different structural domains. *J. Biol. Chem.* **263:** 6751–6758.

Ponten, J., and Saksela, E. (1967). Two established in vitro cell lines from human mesenchymal tumours. *Int. J. Cancer* **2:** 434–447.

Ramaswamy, H., and Hemler, M. E. (1990). Cloning, primary structure and properties of a novel human integrin beta subunit. *EMBO J.* **9:** 1561–1568.

Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G. G., and Stuart, D. (1995). The structure of a Ca(2+)-binding epidermal growth factor-like domain: Its role in protein–protein interactions. *Cell* **82:** 131–141.

Samec, S., Jones, T. A., Corlet, J., Scherly, D., Sheer, D., Wood, R. D., and Clarkson, S. G. (1994). The human gene for xeroderma pigmentosum complementation group G (XPG) maps to 13q33 by fluorescence *in situ* hybridization. *Genomics* **21:** 283–285.

Sheppard, D., Rozzo, C., Starr, L., Quaranta, V., Erle, D. J., and Pytela, R. (1990). Complete amino acid sequence of a novel integrin beta subunit (beta 6) identified in epithelial cells using the polymerase chain reaction. *J. Biol. Chem.* **265:** 11502–11507.

Takagi, J., Isobe, T., Takada, Y., and Saito, Y. (1997). Structural interlock between ligand-binding site and stalk-like region of β1 integrin revealed by a monoclonal antibody recognizing conformation-dependent epitope. *J. Biochem.* **121:** 914–921.

Taylor, M. R., Couto, J. R., Scallan, C. D., Ceriani, R. L., and Peterson, J. A. (1997). Lactadherin (formerly BA46), a membrane-associated glycoprotein expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp (RGD)-dependent cell adhesion. *DNA Cell Biol.* **16:** 861–869.

Timpl, R., Aumailley, M., Gerl, M., Mann, K., Nurcombe, V., Edgar, D., and Deutzmann, R. (1990). Structure and function of the laminin–nidogen complex. *Ann. N.Y. Acad. Sci.* **580:** 311–323.

Tsuji, T., Okada, F., Yamaguchi, K., and Nakamura, T. (1990). Molecular cloning of the large subunit of transforming growth factor type beta masking protein and expression of the mRNA in various rat tissues. *Proc. Natl. Acad. Sci. USA* **87:** 8835–8839.

Wei, Y. F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J., Yu, G. L., Wang, R. P., Shell, B. K., Nash, R. A., Schar, P., *et al.* (1995). Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol. Cell Biol.* **15:** 3206–3216.

Yang-Feng, T. L., Li, S., Han, H., and Schwartz, P. E. (1992). Frequent loss of heterozygosity on chromosomes Xp and 13q in human ovarian cancer. *Int. J. Cancer* **52:** 575–580.

Yin, W., Smiley, E., Germiller, J., Mecham, R. P., Florer, J. B., Wenstrup, R. J., and Bonadio, J. (1995). Isolation of a novel latent transforming growth factor-beta binding protein gene (LTBP-3). *J. Biol. Chem.* **270:** 10147–10160.

Yuan, Q., Jiang, W-M., Krissansen, G. W., and Watson, J. D. (1990). Cloning and sequence analysis of a novel β2-related integrin transcript from T lymphocytes: Homology of integrin cysteine-rich repeats to domain III of laminin B chains. *Int. Immunol.* **2:** 1097–1108.

Yuan, Q., Jiang, W-M., Leung, E., Hollander, D., Watson, J. D., and Krissansen, G. W. (1992). Molecular cloning of the mouse integrin beta 7 subunit. *J. Biol. Chem.* **267:** 7352–7358.

ELSEVIER

Biochimica et Biophysica Acta 1435 (1999) 61–70

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bba

A family of *Arabidopsis* plasma membrane receptors presenting animal β-integrin domains

Valérie Laval [a], Matthieu Chabannes [a], Marguerite Carrière [a], Hervé Canut [a], Annick Barre [b], Pierre Rougé [b], Rafael Pont-Lezica [a], Jean-Philippe Galaud [a,*]

[a] *UMR CNRS/UPS 5546, Pôle de Biotechnologies Végétales, 24, Chemin de Borde-Rouge, BP 17, Auzeville, 31236 Castanet-Tolosan, France*

[b] *Institut de Pharmacologie et Biologie Structurale, UPR CNRS 9062, 205, route de Narbonne, 31077 Toulouse Cédex, France*

Received 16 June 1999; received in revised form 17 August 1999; accepted 18 August 1999

Abstract

A cDNA clone, *AtELP1* (*Arabidopsis thaliana* EGF receptor-like protein) was isolated from an *Arabidopsis* cDNA library with an oligonucleotide probe corresponding to a highly conserved region of animal β-integrins. The cloning of this cDNA was previously reported and it has been proposed that AtELP might be a receptor involved in intracellular trafficking. In the present work, using two specific independent sets of anti-peptide antibodies, we show that AtELP1 is mainly located in the plasma membrane, supporting another function for this protein. Structural studies, using methods for secondary structure prediction, indicated the presence of cysteine-rich domains specific to β-integrins. Database searches revealed that *AtELP1* is a member of a multigenic family composed of at least six members in *A. thaliana*. Northern blot analysis of *AtELP1*, *2b* and *3* was performed on mRNA extracted from cells cultured in normal and stressed conditions, and from several organs and plants submitted to biotic or abiotic stresses. All the genes are expressed at different levels in the same conditions, but preferentially in roots, fruits and leaves in response to water deficit.

Keywords: *Arabidopsis thaliana*; β-Integrin; Cysteine-rich domain; EGF domain; Plasma membrane receptor

1. Introduction

Plant cell morphogenesis is the result of numerous mechanisms involved in the control of cell division and expansion. The cell wall, the plasma membrane, and the cytoskeleton are considered as the main actors in the establishment of polarity and morphogenesis. It is now clear that the membrane and membrane proteins are kept in a dynamic state to maintain cell structure and compartmentalization [1]. Linkages between the plant plasma membrane and the cell wall can be observed after plasmolysis; however, the molecules engaged in this interaction are unknown. In animal cells, integrins are plasma membrane receptors involved in cellular adhesion. Some of them recognize extracellular proteins via the RGD sequence, a conserved motif in adhesion proteins from the extracellular matrix.

The occurrence of integrins in plants has been suggested, but their identification remains obscure. Two lines of evidence support the occurrence of integrin-like receptors in plants. On one hand, immunological cross reactivity between antibodies raised against animal integrins and plant proteins has been observed.

* Corresponding author. Fax: +33-562-193502; E-mail: galaud@cict.fr

0167-4838/99/$ – see front matter
PII: S0005-2728(99)00087-0

Immunological approaches identify several plant proteins sharing common epitopes with animal integrins [2–7]. Using an animal integrin polyclonal antibody for screening an *Arabidopsis* cDNA library, a membrane-associated protein involved in trafficking was isolated [8]. These results show that there are plant proteins sharing some motifs with animal integrins, but the proteins have no homology.

On the other hand, RGD peptides interfere with several plant physiological processes. Indeed, the addition of RGD peptides disrupts protoplast adhesion in tobacco derived from NaCl-adapted cells, [9]. These peptides also inhibit gravity perception in *Chara* [10], and enhance soybean cellular division [2]. Furthermore, in the brown alga *Fucus*, polarity determination is affected by the addition of RGD peptides [11]; and in *Uromyces*, a plant pathogenic fungus, appressorium formation is inhibit by the same compounds [5]. In agreement with these data, *Arabidopsis* plasma membrane exhibits specific high affinity binding sites for RGD-containing peptides and proteins. RGD binding is strongly inhibited by trypsin treatment, supporting the protein nature of the receptor [12].

In this paper, we use an oligonucleotide screening strategy to clone integrin-like molecules in *Arabidopsis*. The oligonucleotide probe is defined according to a cytoplasmic conserved region of integrin β-subunit, involved in interactions with the cytoskeleton [13]. If homology between animal and plant integrins exists, a functional domain involved in the interaction with cytoskeleton proteins will be present. Little homology is expected at the extracellular level since the extracellular matrices of animals and plants are completely different. The receptors binding to these matrices should reflect these differences [14].

2. Materials and methods

2.1. Plant material

Arabidopsis thaliana, ecotype Columbia, was cultured in a grown chamber under fluorescent tubes 36 W (12 W/m^2) with 16-h light–8-h dark photoperiod. Plants were grown in pots filled with TKS2 peat Floratorf supplemented with 1‰ (w/w) nitrate. *Arabidopsis* cells were grown on Gamborg liquid medium [15]. Fifteen ml cell suspensions were routinely transferred to 300 ml fresh medium in 1000-ml Erlenmeyer flasks every 2 weeks, and shaken (150 rpm) in continuous light (60 W/m^2) at 26°C. Cells were transferred to fresh culture medium containing mannitol or not (250 mM), and maintained in the dark for different periods (1–15 days) prior to harvesting.

2.2. cDNA library screening, sequencing and computer sequences analysis

The *A. thaliana* cDNA library was constructed in pAD-GAL4 vector (Stratagene) and was kindly provided by B. Lescure (INRA-CNRS, Auzeville). A 17-mer degenerate oligonucleotide (AARTTYGAR-AARGARAA) corresponding to the peptide sequence KFEKEK was synthesized. This peptide matched a cytoplasmic conserved region from human integrin (β-subunit) [16]. The oligonucleotide was labeled with [γ-^{32}P]ATP by terminal transferase and used to screen the cDNA library according to Stratagene protocol. Positive clones were selected, excised from recombinant phage, and introduced into *Escherichia coli* strain SOLR. The isolated cDNAs were sequenced according to Sanger et al. [17].

The DNA and its deduced protein sequences were examined for homology in the non-redundant nucleotide and protein sequence databases using BLAST [18], PRODOM [19], and BLOCKS searches [20]. The amino acid sequence alignments were carried out on a Macintosh LC630 computer. The hydrophobicity, surface probability and flexibility profiles were calculated as described [21–23] with a window size of seven residues, using MacVector (Kodak). Hydrophobic cluster analysis (HCA) [24,25] was performed to delineate and compare the hydrophobic clusters along the amino acid sequences. They were generated on a Macintosh LC using the program HCA-Plot2 (Doriane, Paris, France).

2.3. RNA isolation and DNA–RNA hybridization analysis

Total RNA was extracted from *Arabidopsis* cell suspensions at various times during the culture and from different organs using the guanidinium thiocyanate method [26]. Total RNA (15 μg) was separated on formaldehyde agarose gel and blotted to Nytran

(Schleicher and Schuell) according to the manufacturer's specifications. The 3′-non-coding region of *AtELP1*, *2*, and *3* were amplified by PCR using degenerate primers, deduced from the highly conserved region of the three clones (5′-ATCATGKCACAGTAYATGCCA-3′) and a TTTTTTTTTTTTTTTW. Each PCR fragment was subcloned in pGEM-T vector (Promega), sequenced and used as specific probe.

2.4. Antibody production and purification

Specific antibodies raised against selected peptides derived from AtELP1 were prepared. Immunogenic peptides were defined by HCA and prediction of antigen determinants [27]. Two exposed hydrophilic regions, located between amino acids 352 and 365 (AEQESQIGKSRGDC, peptide 63), and amino acids 375–384 (NNRQYRGKLEC, peptide 64), were defined. BLAST analysis was carried-out to verify the presence of identical sequences in the *Arabidopsis* database. No other known protein, but the AtELP family, showed sequence 63 or 64 indicating that the chosen peptides could be specific for AtELP proteins. Both peptides were synthesized automatically by stepwise F-moc-*t*-butyl solid phase synthesis [28] in a Synergy Applied Biosystems peptide synthesizer. Crude synthetic peptides were purified by reverse-phase HPLC. Purified peptides were characterized by mass spectrometry on a Lasermat spectrometer (Finnigan), and coupled to the carrier protein. Peptides were coupled either to tyroglobulin or to bovine serum albumin using *N*-succimidyl-6-maleidocaproate as coupling reagent.

Before immunization, a sample of preimmune serum was taken and tested against peptides 63 and 64. In the absence of response, the immunization was performed. One volume of complete (immunization) Freund's adjuvant was added to the tyroglobulin-coupled peptide (250 μg per injection) and injected into rabbits. Two rabbits were immunized against each coupled peptide every 2 weeks during 3 months. Two antisera were obtained: serum 630 for peptide 63, and serum 640 for peptide 64. Antibodies were immunopurified before use. Ten micrograms of BSA-coupled peptide was separated by SDS-PAGE and transferred to nitrocellulose. The membrane was stained with Ponceau red. The stained region was cut, unstained, and blocked with TBS, 0.1% Tween, and 10% non-fat milk for 1 h at room temperature. The membrane was washed three times (15 min each) with TBS 0.1% Tween 1% BSA, and incubate overnight at 4°C with (1/25 dilution) serum. The antibodies were eluted with 500 μl glycine EGTA buffer (glycine 0.2 M, EGTA 1 mM, pH 2.8) and neutralized with 70 μl Tris 1 M pH 8.

2.5. Fractionation of A. thaliana membranes

Microsomes from *Arabidopsis* cells were prepared according to Bardy et al. [29] with a grinding medium containing 0.17 M sucrose, 50 mM KCl, 1 mM DTT, and 10 mM HEPES, pH 7.5. Microsomes were separated by free-flow electrophoresis with an Elphor Vap-22 electrophoresis unit (Weber, Kirchheim-Heimstetten, Germany). The electrophoresis medium contained 0.25 M sucrose, 10 mM KCl, 1 mM $MgCl_2$, 10 mM Tris and 10 mM boric acid (pH 8.3). The electrode buffer consisted of 100 mM Tris, 100 mM boric acid (pH 8.3). Microsomes were resuspended in electrophoresis medium and centrifuged for 30 min at $45\,000 \times g$. Electrophoresis was performed at a 100 mA constant current (about 900 V), sample injection 2 ml h^{-1}, and buffer flow 3.5 ml $fraction^{-1}$ h^{-1} at 4°C. The distribution of membranes in each separation was monitored by absorbance at 280 nm. Membranes were collected from pooled fractions by centrifugation (30 min at $45\,000 \times g$). Activity of different marker enzymes was determined as previously described [29]. Protein content was determined as reported [30] with bovine serum albumin as standard.

2.6. Gel electrophoresis and immunodetection

Gel electrophoresis was carried-out on 11% acrylamide gels. Samples (50 μg purified protein) were solubilized in 0.125 M Tris pH 6.8, 4% SDS and 20% glycerol prior to electrophoresis. Proteins were transferred to nitrocellulose, and incubated overnight with 630 or 640 (1/100 dilution) purified primary antibodies, washed, and revealed with ImmunoPure ABC phosphatase staining kit (Pierce).

Antibody competition was realized by incubation of 1 mg non-coupled peptide with its corresponding antibody for 2 h at 37°C. The exhausted antibody

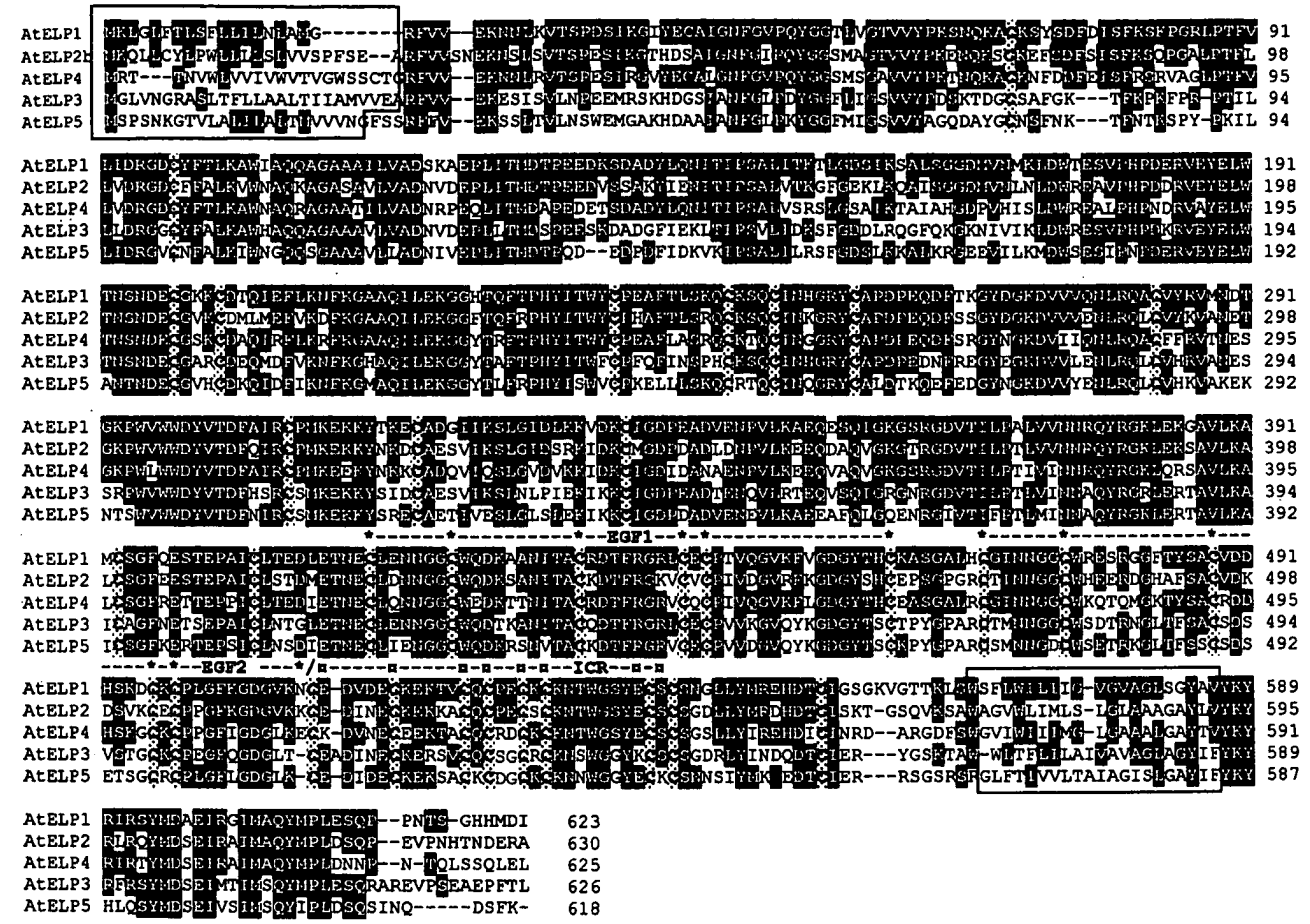


Fig. 1. Comparison of the predicted protein sequences of AtELP1, 2b, 3, 4, and 5. Residues identical in all the proteins are highlighted and the conserved cysteine residues are in gray. EGF signature (epidermal growth factor) is shown by * and ICR (integrin cysteine-rich motif) by a boxed circle. Putative peptide signal and transmembrane domains are boxed. Spaces, denoted by dashes have been introduced to optimize the alignment.

was then incubated with the nitrocellulose membranes.

3. Results

3.1. Molecular cloning, homology searches, and protein sequence analysis

An *A. thaliana* cDNA library was screened with an oligonucleotide probe corresponding to a conserved cytoplasmic region of integrin β-subunits. Seven clones were isolated and partially sequenced. One clone (2712), presenting a potential transmembrane domain, was completely sequenced. The cDNA insert was found to be 2314 bp in length. It encodes a complete 623 amino acid protein, which has a predicted molecular mass of 70 kDa and a potential membrane-spanning domain. Database search revealed that this clone was independently identified by three other groups at the same time [31]. Clone *2712* will be called *AtELP1* (*A. thaliana* EGF-like protein) in this paper.

Southern blot experiments using *AtELP1* cDNA as a probe (data not shown) revealed that *AtELP1* belongs to a multigenic family. Homology searches in *A. thaliana* EST database showed two nucleic acid sequences having 81 and 63% homology with *AtELP1*. These sequences were called, respectively, *AtELP2* (accession number U79960) and *AtELP3* (EST accession number 110G6T7). The complete genomic sequence of *AtELP3* (accession number

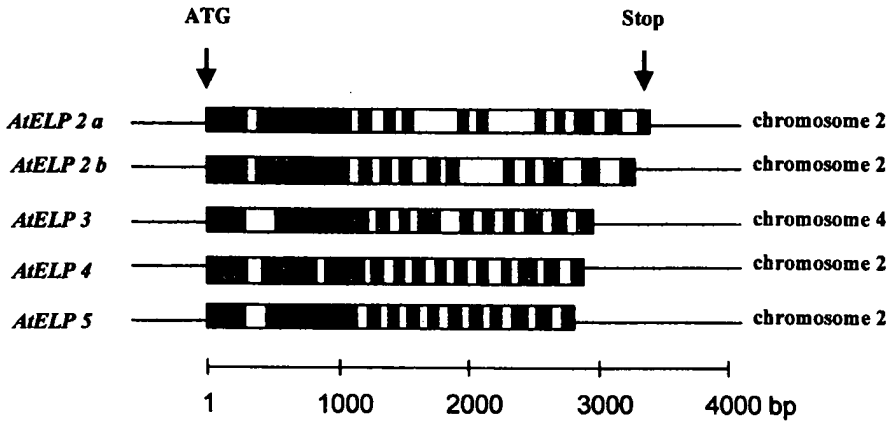


Fig. 2. Schematic representation of *AtELP2a*, *2b*, *3*, *4*, and *5* genomic clones. *AtELP2a*, *2b* (BAC F26C24), *4* and *5* (BAC T09D09, and BAC F19I3) are located on chromosome 2. *AtELP3* (BAC F18F4) is located on chromosome 3. Exons are represented by gray boxes and introns by white boxes. ATG represent the translation start codon.

AL021637, pid g 2827665) and two other clones presenting homologies with *AtELP1* were deduced from the analysis of the *A. thaliana* genomic database. These clones were named *AtELP4* (accession number ATAC002338, pid g 2347209) and *AtELP5* (accession number ATAC004238, pid g 3033390) and their nucleic acid sequences showed 84 and 61% homology with *AtELP1*.

Primary and secondary sequence analyses were performed on AtELP1. The polypeptide chain is rather hydrophilic (64% polar residues) and contains a high proportion of cysteine (34 residues). Its calculated isoelectric point is 5.87. The polypeptide shares moderate percentages of identity (14.8 and 17.8%) and homology (34.6 and 38.8%) with human β1 and β5 integrins.

The deduced protein sequences of the five clones are shown in Fig. 1. AtELP proteins ranged from 618 to 630 amino acids. They present the same common structural features: a potential signal peptide at the N-terminus, a large N-terminal region, a potential transmembrane domain, and a short C-terminal region. Alignment of AtELP1, 2, and 4 showed about 80% homology. The N-terminus contains many conserved regions. The number and position of cysteines is well conserved (34 Cys out of 560 amino acids). Three Cys-rich motifs were found in all the putative proteins. Two of them (EGF1 and EGF2, in Fig. 1) have the typical arrangement $Cx_{(3-7)}Cx_{(2-6)}Cx_{(7-10)}CxCx_{(7-12)}C$ common to epidermal growth factors (EGF) with 6 cysteines in conserved positions. The third Cys-rich motif has a different organization with eight cysteines in the following sequence: $Cx_{(6-8)}Cx_5CxCxxCxCx_{(8-13)}Cx_{(1-2)}C$. This cysteine alignment is characteristic of β-integrin subunits (ICR, integrin Cys-rich motif in Fig. 1). The C-terminal domain of AtELPs (34–40 amino acids) contains a highly conserved sequence of 27 amino acids, but then diverges. It contains a YMPL site (amino acid 606–609). The Yxxϕ motif (x represents any amino acid and ϕ a hydrophobic residue) has been demonstrated to mediate internalization from the cell surface as well as targeting to intracellular compartments in mammals [32]. The more divergent sequences correspond to the potential signal peptide, the putative transmembrane domain, and the C-terminus end.

3.2. Structure of genomic clones

The structure of five *AtELP* genes is presented in Fig. 2. These genomic sequences were obtained by the systematic sequencing programs [33]. Two sequences, *AtELP2a* (accession number ATAC 004705, pid 3252813) and *AtELP2b* (accession number ATAC 004705, pid 3252815, which corresponds to the cDNA previously described as *AtELP2*), encode proteins showing 96.5% identity (22 different amino acids out of 628). The length of the ORF is the same for both genes, but they have different intron lengths. *AtELP2a* and *AtELP2b* are located on

chromosome 2 in reversed position, and are separated by a single gene. *AtELP2a* and *2b* could be the result of gene duplication. *AtELP4* (accession number ATAC 002338, pid 2347209) and *AtELP5* (accession number ATAC 004238, pid 3033390) are

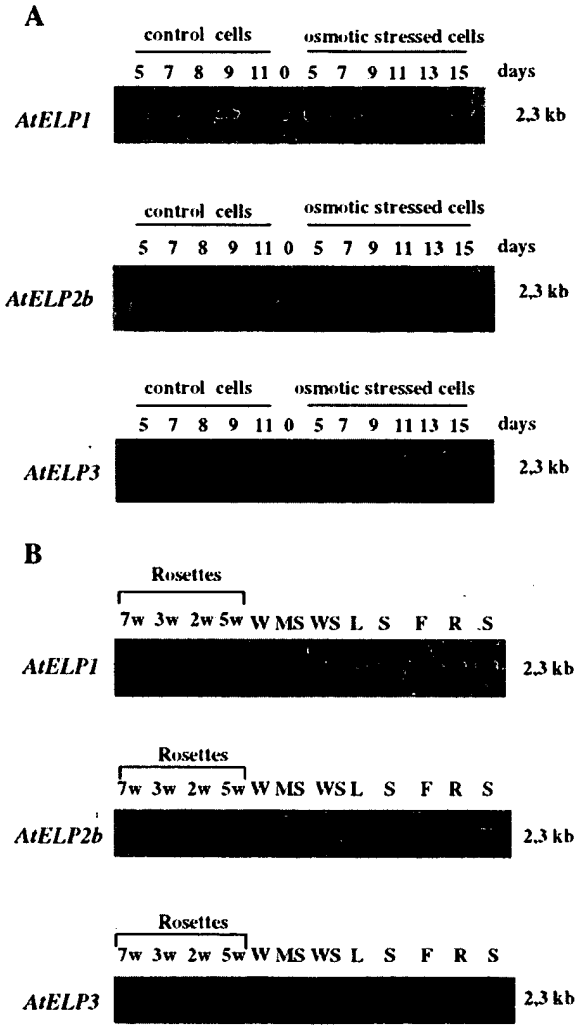


Fig. 3. (A) Northern blot analysis of *AtELP1, 2b*, and *3* gene expression during cell culture in control and osmotic stress conditions. Total RNA was extracted from cells cultured with 250 mM mannitol (osmotic stress) or not (control cells). Total RNA (15 μg) was separated on formaldehyde gel blotted and probed with ^{32}P-specific 3′-UTR from each clone. (B) Northern blot analysis of *AtELP1, 2b*, and *3* gene expression in various organs of *Arabidopsis*. Total RNA was extracted from: leaves (L), stem (S), flowers (F), roots (R), siliques (S) and rosettes at different developmental stages (7w, 7 weeks; 3w, 3 weeks; 2w, 2 weeks; 5w, 5 weeks) and after several stresses (W, wounding; MS, mechanical stress; WS, water stress).

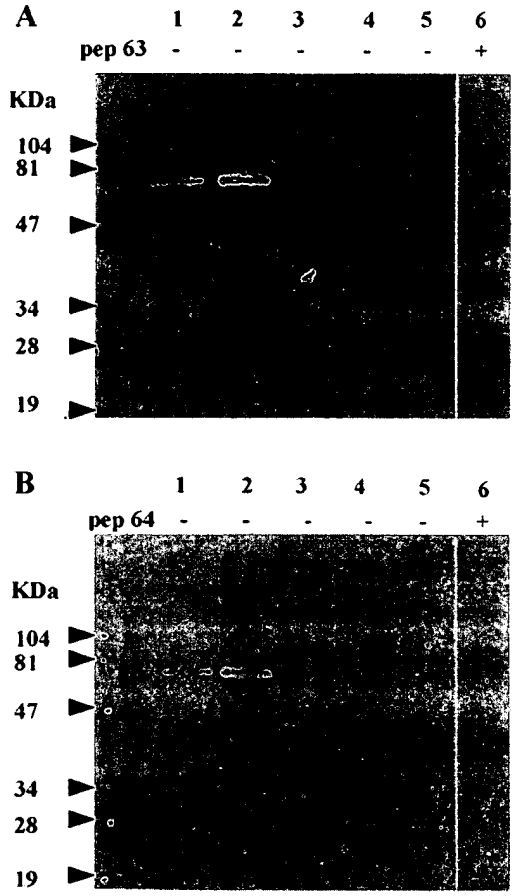


Fig. 4. Western blot analysis of AtELP1 on purified membrane fractions by free-flow electrophoresis. Membranes were prepared from cells cultured in mannitol containing medium. Fractions number 1 and 2 correspond to plasma membrane, fractions 3 and 4 to endomembrane (Golgi, ER, mitochondria...) and fraction 5 to tonoplast. Lane 6 corresponds to plasmalemma (lane 2). Fifty μg protein were separated on by 11% SDS PAGE gel, blotted onto nitrocellulose and incubated with antibody 630 (A) and antibody 640 (B) in the presence (+) or in absence (−) of the corresponding peptide. Molecular mass standards are given on the left of the figure.

also located on chromosome 2. The genomic sequence corresponding to *AtELP3* (accession number ALO21637, pid 2827665) is the only one located on chromosome 4. At present, the *AtELP1* genomic clone has not been found. The multigenic family *AtELP* is composed of at least six genes having 11–13 exons and 10–12 introns. Among these six genes, only three have so far been found expressed (*AtELP1*, *AtELP2b*, and *AtELP3*).

Sequence analysis of each promoter was carried out using the Transfac program [34]. No clearly iden-

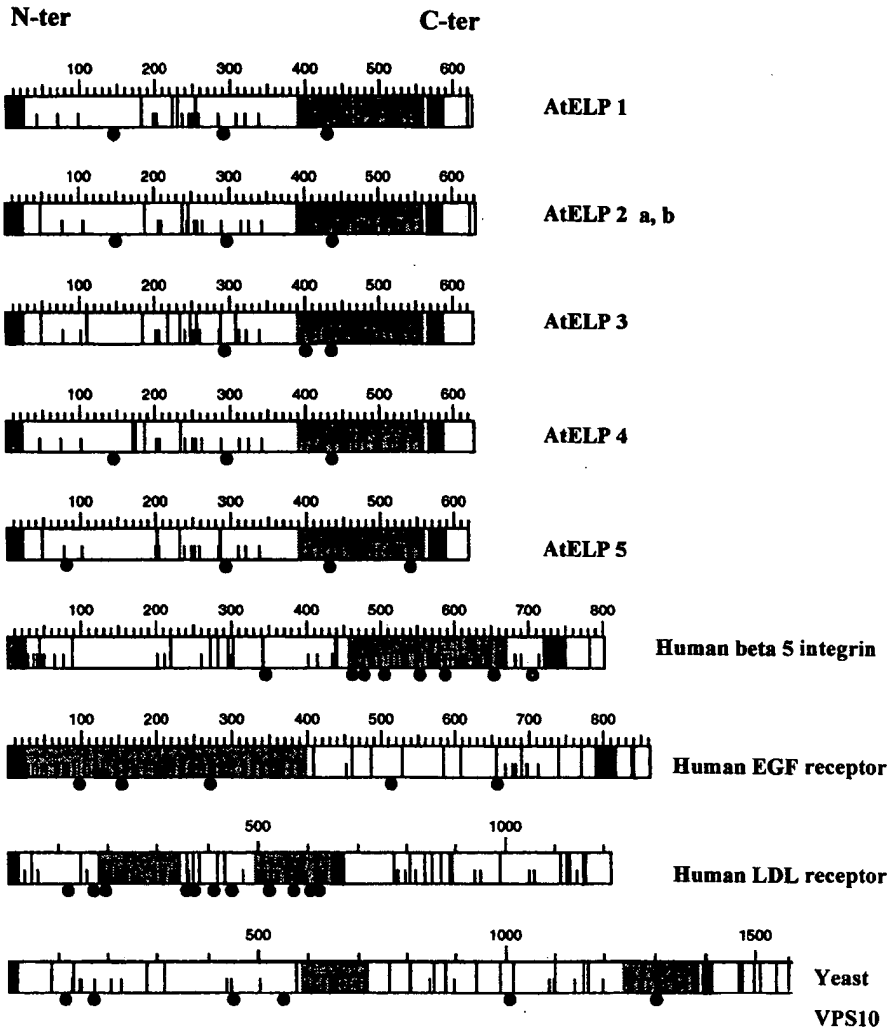


Fig. 5. Schematic comparison of AtELP1, 2a, 2b, 3, 4, 5 with other cysteine-rich proteins. Human β5-integrin subunit (accession number M35011), human EGF receptor (accession number X00588), human LDL receptor (accession number L00352) and the yeast VPS10 (accession number U07621). Cysteine-rich domains are represented by gray boxes and cysteines by small bars. At the N-ter end, peptide signal is represented by hatched boxes; transmembrane domains are black boxed and black circles represent putative glycosylation sites.

tifiable regions corresponding to putative *cis*-acting elements were found in *AtELP* promoter genes.

3.3. *AtELP1, AtELP2b and AtELP3 gene expression analysis*

Analysis of *AtELP1* gene expression was done using the 3′ non-coding region, cloned after PCR amplification. Total RNA was extracted from dark cultured cells under control or osmotic stress (Fig. 3A). A single signal corresponding to a 2.3-kb transcript was observed. This signal was increased during the culture period under osmotic stress, compared to control cells. In Fig. 3B, *AtELP1* gene expression analysis was performed on total RNA extracted from different organs or after various stress: mechanical, wounding, water deficit, and after *Ralstonia solanacearum* infection (data not shown). A weak signal was observed in young plantlets (rosette stage) compared to other organs (leaf, root, and stem). Stress was applied on plantlets at 5 weeks rosette stage, and the higher signal was observed in plants

left without water for 24 h. *AtELP2b* and *AtELP3* gene expression was analyzed using specific probes, 3′-non-coding region were specifically PCR amplified before use. Expression of *AtELP2b* is weakly enhanced compared to *AtELP1* (Fig. 3A,B) and the basal expression level of *AtELP3* is weaker than the other genes under the conditions used.

3.4. AtELP1 localization

Two sets of antibodies (630 and 640) raised against peptides derived from the AtELP1 primary sequences were used for the localization of the protein. Purified membranes were obtained and characterized as described [29]. The proteins were separated on SDS gel electrophoresis, transferred to nitrocellulose membranes, and revealed using antibody 630. A single band around 80 kDa was observed in plasma membrane and endomembrane enriched fractions (Fig. 4A). The higher signal was observed in fraction 2 (plasma membrane). This signal strongly decreased when antibody 630 was pre-incubated with its corresponding peptide (lane 6). The same results were obtained when antibody 640 was used (Fig. 4B).

4. Discussion

A new class of membrane proteins (AtELPs) has been identified. Database search revealed that *AtELP1* was simultaneously cloned by three other groups. Whereas the primary sequence of the protein does not show obvious homology with known proteins from plants or other organisms, it has been proposed that AtELP might be a receptor involved in plant intracellular protein trafficking [35,36]. AtELP1's secondary structure was compared with that of two well-known trafficking sorting proteins: the yeast VSP10 and the rat mannose-6-phosphate receptor (M6PR). Dot-plots performed with the PAM250 matrix [37] showed a limited number of very short diagonals. This indicated that no consistent similarities occurred between the proteins (data not shown). In addition, the Cys-rich domains located close to the C-terminal of AtELP1 do not occur in other proteins.

HCA analysis of AtELP1 and human integrins β1 and β5 reinforced the structural comparison indicating that these three proteins exhibit a very similar molecular organization characterized by the Cys-rich domains. AtELP1 and β-integrin subunits have 13.1 and 15% cysteine, respectively, on the stretch preceding the transmembrane domain [38]. These cysteines are presumably disulfide-bonded and such bonding would necessarily occur in the extracellular domain [39]. As in the β-integrin family, AtELPs contain two EGF-like signatures. The third Cys-rich domain present in AtELPs is characteristic of β-integrins as indicate in the PROSITE database. Such Cys-rich repeats seem to stabilize the integrin structure at the base of the protein.

EGF domains have been found in a large number of proteins and their common feature is to be present in the extracellular domain of membrane or secreted proteins, with the exception of a prostaglandin G/H synthase. In SWISS-PROT database, 49 proteins present the EGF signature, containing six cysteines. In the mammalian EGF and LDL receptors, EGF regions seem to be involved in receptor–ligand interactions at the cell surface of animal cells [40,41].

Fig. 5 compares AtELP1–5 with other Cys-rich membrane proteins, like the LDL receptor, the EGF receptor, the yeast Vps10, and the human integrin β5. From a structural point of view, AtELPs seem to be closer to the integrin family than to the endocytic or sorting receptors. AtELPs and the β-subunit of animal integrins may be derived from a common ancestor. AtELPs may be considered as integrin orthologous proteins. Brower et al. [42] suggest that cell surface receptors are strongly conserved in higher animals, but their molecular evolution remains obscure. The cloning of two cDNAs encoding integrin β-subunits from coral and sponge clearly showed that the major structural features were well conserved. Comparative analysis of the genome of the nematode *Caenorhabditis elegans* showed that the main differences observed between the two genomes corresponded to proteins involved in the establishment of multicellularity (adhesion molecules) and cell death machinery (signaling proteins) [43]. Brower et al. [42] indicated that even if no protein with obvious homology to integrins has been identified, the existence of integrin-like molecules in plants [4,12] and in fungi [44,45] has been reported by several authors.

Integrins recognize the RGD sequence in their li-

gand via interaction with DxSxS, an integrin binding motif [46]. The presence of an equivalent motif is observed in AtELP1, 2a, and 2b proteins at the N-terminal domain. The presence on the same polypeptide of a motif binding RGD (DxSxS), and an exposed RGD sequence is puzzling. A RGD sequence in the N-terminus is also present in the extracellular domain of human β2, β5 and β6 integrins. Papadopoulos et al. [47] reported that 7182 proteins contain the RGD motif. Of these proteins, only 120 are membrane or membrane-associated proteins having the RGD sequence in their extracellular domain and some are proteins involved in cell adhesion processes.

Ahmed et al. [36] showed that the C-terminal region of AtELP1 is located in the cytosol. The cytoplasmic domain of AtELPs is well conserved and it contains the Yxxϕ motif at a distance of about 20 amino acids from the membrane-spanning region. This sequence is a recognition motif for endocytosis through clathrin-coated vesicles [32]. In animal integrins, the presence of the signal sequence (NPxY), localized about 20–25 amino acids from the transmembrane domain [16], is required for internalization via clathrin-coated vesicles [48]. In this line, the vitronectin receptor αvβ5 plays a double role in fibroblasts; it binds to and directly internalizes vitronectin. The presence of AtELP on the plasma membrane indicates that the protein follows the secretory pathway when it is synthesized, and may be internalized via clathrin-coated vesicles into the vacuolar compartment. This is supported by the presence of the protein in the trafficking vesicles [35], as well as for the presence of internalization signals in the cytoplasmic tail.

At present, six genes encoding AtELP were identified in *Arabidopsis*, but only three were expressed, referred to EST and cDNA sequencing programs. The main accumulation of transcripts was observed in roots, in cultured cells and in young plantlets submitted to water stress. Previous work from Katembe et al. [6], showed that integrin-like molecules were accumulated in *Arabidopsis* roots, an important gravity perception site. Wayne et al. [10] reported that gravisensing of *Chara* cells was dependent on RGD binding and a collagenase treatment indicated the presence of a PxGP motif in the RGD gravireceptor sequence. This motif is present in a conserved position in AtELP2a, 2b, 3, and 5 in the large N-terminal region, but the biological significance is unknown.

Acknowledgements

V.L. is a fellow from the Ministere de l'Education Nationale, de la Recherche et de la Technologie, France. We are especially grateful to N. Vita (Sanofi-Recherche, Labège) and H. Mazarguil (IPBS, Toulouse) for their helpful collaboration, and to the *Arabidopsis* Biological resource Center, Ohio State University, for providing EST clones. 'Merci à Lorena' for English revision. This project was supported by Université Paul Sabatier and by the Centre National de la Recherche Scientifique (Grants 96N88/0010 and 285016).

References

[1] J.E. Fowler, R.S. Quatrano, Annu. Rev. Cell Dev. Biol. 13 (1997) 697–743.
[2] M. Schindler, S. Meiners, D.A. Cheresh, J. Cell Biol. 108 (1989) 1955–1965.
[3] R.S. Quatrano, L. Brian, J. Aldridge, T. Schultz, Development 1 (Sup.) (1991) 11–16.
[4] J.S. Gens, C. Reuzeau, K. Doolittle, J.G. McNally, B.G. Pickard, Protoplasma 194 (1996) 215–230.
[5] A. Corrêa, R.C. Staples, H.C. Hoch, Protoplasma 194 (1996) 91–102.
[6] W.J. Katembe, L.J. Swatzell, C.A. Makaroff, J.K. Kiss, Physiol. Plant. 99 (1997) 7–14.
[7] A. Faik, A.M. Labouré, D. Gulino, P. Mandaron, D. Falconet, Eur. J. Biochem. 253 (1998) 552–559.
[8] J.P. Galaud, V. Laval, M. Carriere, A. Barre, H. Canut, P. Rouge, R. Pont-Lezica, Biochim. Biophys. Acta 1341 (1997) 79–86.
[9] J.K. Zhu, J. Shi, U. Singh, S.E. Wyatt, R.A. Bressan, P.M. Hasegawa, N.C. Carpita, Plant J. 3 (1993) 637–646.
[10] R. Wayne, M.P. Staves, A.C. Leopold, J. Cell Sci. 101 (1992) 611–623.
[11] D.L. Kropf, B. Kloareg, R.S. Quatrano, Science 239 (1988) 187–190.
[12] H. Canut, A. Carrasco, J.P. Galaud, C. Cassan, H. Bouyssou, N. Vita, P. Ferrara, R. Pont-Lezica, Plant J. 16 (1998) 63–71.
[13] M. Pfaff, S. Liu, D.J. Erle, M.H. Ginsberg, J. Biol. Chem. 273 (1998) 6104–6109.
[14] C. Reuzeau, R.F. Pont-Lezica, Protoplasma 152 (1995) 113–121.

[15] M. Axelos, C. Curie, L. Mazzolini, C. Bardet, B. Lescure, Plant Physiol. Biochem. 30 (1992) 123–128.
[16] S. K Sastry, A.F. Horwitz, Curr. Opin. Cell Biol. 5 (1993) 819–831.
[17] F. Sanger, S. Nicklens, A.R. Coulson, Proc. Natl. Acad. Sci. USA 74 (1997) 5463–5467.
[18] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D. Lipman, J. Mol. Biol. 215 (1990) 403–410.
[19] E.L. Sonnhammer, D. Kahn, Protein Sci. 3 (1994) 482–492.
[20] S. Henikoff, J.G. Henikoff, Genomics 19 (1994) 97–107.
[21] J. Kyte, R.F. Doolittle, J. Mol. Biol. 157 (1982) 105–132.
[22] J. Janin, Nature 277 (1979) 491–492.
[23] P.A. Karplus, G.E. Schulz, Naturwissenschaften 72 (1985) 212.
[24] C. Gaboriaud, V. Bissery, T. Benchetrit, J.P. Mornon, FEBS Lett. 224 (1987) 149–155.
[25] L. Lemesle-Varloot, B. Henrissat, C. Gaboriaud, V. Bissery, A. Morgat, J.P. Mornon, Biochimie 72 (1990) 555–574.
[26] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: a Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[27] T.P. Hopp, K.R. Woods, Proc. Natl. Acad. Sci. USA 78 (1981) 3824–3828.
[28] G.B. Fields, R.L. Noble, Int. J. Pept. Protein Res. 35 (1990) 161–214.
[29] N. Bardy, A. Carrasco, J.P. Galaud, R. Pont-Lezica, H. Canut, Electrophoresis 19 (1998) 1145–1153.
[30] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Anal. Biochem. 150 (1985) 76–85.
[31] J.-M. Neuhaus, J.C. Rogers, Plant Mol. Biol. 38 (1998) 127–144.
[32] H. Ohno, J. Stewart, M.-C. Fournier, H. Bosshart, I. Ree, S. Miyatake, T. Saito, A. Gallusser, T. Kirchhausen, J.S. Bonifaciono, Science 269 (1995) 1872–1875.
[33] M.L. Wang, L. Huang, D.K. Bongard-Pierce, S. Belmonte, E.A. Zachgo, J.W. Morris, M. Dolan, H.M. Goodman, Plant J. 12 (1997) 711–730.
[34] T. Heinemeyer, E. Wingender, I. Reuter, H. Hermjakob, A.E. Kel, O.V. Kel, E.V. Ignatieva, E.A. Ananko, O.A. Podkolodnaya, F.A. Kolpakov, N.L. Podkolodny, N.A. Kolchanov, Nucleic Acids Res. 26 (1998) 362–367.
[35] N. Paris, S.W. Rogers, L. Jian, T. Kirsch, L. Beevers, T.E. Phillips, J.C. Rogers, Plant Physiol. 115 (1997) 29–39.
[36] S.V. Ahmed, M. Bar-Peled, N.V. Raikhel, Plant Physiol. 114 (1997) 325–336.
[37] M.O. Dayhoff, R.V. ECK, C.M. Park, in: Atlas of Protein Sequence and Structure, Vol. 5, Natl. Biomed. Foundation, Washington, DC, USA, 1972.
[38] J.W. Tamkun, D.W. DeSimon, D. Fonda, R.S. Patel, C. Buck, A.F. Horwitz, R.O. Hynes, Cell 46 (1986) 271–282.
[39] J.J. Calvete, A. Henschen, J. Gonzalez-Rodriguez, Biochem. J. 274 (1991) 63–171.
[40] T. Yamamoto, C.G. Davis, M.S. Brown, W.J. Schneider, M.L. Casey, J.L. Goldstein, D.W. Russell, Cell 39 (1984) 27–38.
[41] A. Lalazar, K.H. Weisgraber, S.C. Ra, H. Giladi, T.K.L. Inneraarity, A.Z. Levanon, J.K. Boyles, B. Amit, M. Gorecki, R.W. Mahley, J. Biol. Chem. 263 (1988) 3542–3545.
[42] D.L. Brower, S.M. Brower, D.C. Hayward, E.E. Ball, Proc. Natl. Acad. Sci. USA 94 (1997) 9182–9187.
[43] S.A. Chervitz, L. Aravind, G. Sherlock, C.A. Ball, E.V. Koonin, S.S. Dwight, M.A. Harris, K. Dolinski, S. Mohr, T. Smith, S. Weng, J.M. Cherry, D. Botstein, Science 282 (1998) 2022–2028.
[44] C. Gale, D. Finkel, N. Tao, M. Meinke, M. McClellan, J. Olson, K. Kendrick, M. Hostetter, Proc. Natl. Acad. Sci. USA 93 (1996) 357–361.
[45] C.A. Gale, C.M. Bendel, M. McClellan, M. Hauser, J.M. Becker, J. Berman, M.K. Hostetter, Science 279 (1998) 1355–1358.
[46] M.L. Bajt, J.C. Loftus, J. Biol. Chem. 269 (1994) 20913–20919.
[47] G.K. Papadopoulos, C. Ouzoumis, E. Eliopoulos, Int. J. Biol. Macromol. 22 (1998) 51–57.
[48] W.J. Chen, J.L. Goldstein, M.S. Browns, J. Biol. Chem. 265 (1990) 3116–3123.

holder Ns) that have no BLAST hit at 99% identity for finished data and 95% identity for light-shotgun data were considered uncovered. The percentage of each clone not hit by Celera sequence was calculated by dividing the total length of the uncovered sequence by the sequence length of the clone. The total number of nucleotides that have no coverage in the Celera assembled contigs was calculated by summing the regions of no hits for all the clones that covered Celera contigs by less than 90% (95% for finished clones). This cutoff value was chosen to eliminate the occasionally low quality of sequences in the clone sequence data. The cutoff value of 90% was determined by the amount of no-hit sequences in 16 light-shotgun clones that are fully contained within three Celera contigs. A higher cutoff value (95%) was used for the finished data than for the light-shotgun data, because finished clones have better sequence quality. The total amount of uncovered sequence for each light-shotgun clones was calculated by multiplying the no-hit percentage of the clone by the clone length as determined by sizing on agarose gels (*36*). For those light-shotgun clones with unreported insert sizes, the sequence length, excluding Ns, was used instead. For finished clones, the amount of uncovered sequence was calculated by multiplying the no-hit percent of the clone by the clone's length. We created 7-kbp subcontig blocks and considered each block to be fully present in the draft sequence if it was hit by at least 500 bp of external sequences. We chose these parameters conservatively, based on the fact that at 1× sequence coverage, the chance of failing to sample a 7-kbp region covered by a light-shotgun clone is 1 in 10^6. For the WGS assembly, we identified 1380 blocks that were hit by less than 500 bp of clone sequence and 794 blocks that were completely missed by the clone sequence. The total number of missed blocks is 2174, which represents a total 15.2 Mbp.

34. M. Ashburner *et al.*, *Genetics* **153**, 179 (1999).
35. Seven conflicts were identified in this study, six of which appear to be owing to transposable elements. The remaining represents a 30-kbp insert within a Celera contig that does not match the corresponding clone. This discrepancy is still under investigation.
36. www.sciencemag.org/feature/data/1049666.shl
37. S. Altschul *et al.*, *Nucleic Acids. Res.* **25**, 3389 (1997).
38. R. A. Hoskins, personal communication.
39. In order to align the Celera sequences unambiguously to the external data, all significant HSPs at the parameters given in (*27*) were screened to identify "mutually unique regions" where the clone and contig sequences have a unique, reciprocal match relation.
40. Most negative gaps arise because of inaccuracies in the distances implied by bundles—the bundle implies a small amount of overlap between two contigs because it is actually short, whereas the reality is that there is a small gap at that location. In a very small number of cases, there is an overlap, but it is because the distance estimate is too long by 3 standard deviations, or because there is a small bit of foreign DNA at the tip of a contig because of untrimmed vector or a chimeric read. None of these negative gaps has yet been found to imply incorrect assembly.

41. We wish to thank H. Smith and S. Salzburg for the many collegial exchanges, M. Peterson and his team for keeping the machines humming, R. Thompson and his staff for providing us with an environment conducive to such an intense effort, and A. Glodek, C. Kraft, and A. Deslattes Mays, and their staff for getting the data to us.

REVIEW

Comparative Genomics of the Eukaryotes

Gerald M. Rubin,[1] Mark D. Yandell,[3] Jennifer R. Wortman,[3] George L. Gabor Miklos,[4] Catherine R. Nelson,[2] Iswar K. Hariharan,[5] Mark E. Fortini,[6] Peter W. Li,[3] Rolf Apweiler,[7] Wolfgang Fleischmann,[7] J. Michael Cherry,[8] Steven Henikoff,[9] Marian P. Skupski,[3] Sima Misra,[2] Michael Ashburner,[7] Ewan Birney,[7] Mark S. Boguski,[10] Thomas Brody,[11] Peter Brokstein,[2] Susan E. Celniker,[12] Stephen A. Chervitz,[13] David Coates,[14] Anibal Cravchik,[3] Andrei Gabrielian,[3] Richard F. Galle,[12] William M. Gelbart,[15] Reed A. George,[12] Lawrence S. B. Goldstein,[16] Fangcheng Gong,[3] Ping Guan,[3] Nomi L. Harris,[12] Bruce A. Hay,[17] Roger A. Hoskins,[12] Jiayin Li,[3] Zhenya Li,[3] Richard O. Hynes,[18] S. J. M. Jones,[19] Peter M. Kuehl,[20] Bruno Lemaitre,[21] J. Troy Littleton,[22] Deborah K. Morrison,[23] Chris Mungall,[12] Patrick H. O'Farrell,[24] Oxana K. Pickeral,[10] Chris Shue,[3] Leslie B. Vosshall,[25] Jiong Zhang,[10] Qi Zhao,[3] Xiangqun H. Zheng,[3] Fei Zhong,[3] Wenyan Zhong,[3] Richard Gibbs,[26] J. Craig Venter,[3] Mark D. Adams,[3] Suzanna Lewis[2]

A comparative analysis of the genomes of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*—and the proteins they are predicted to encode—was undertaken in the context of cellular, developmental, and evolutionary processes. The nonredundant protein sets of flies and worms are similar in size and are only twice that of yeast, but different gene families are expanded in each genome, and the multidomain proteins and signaling pathways of the fly and worm are far more complex than those of yeast. The fly has orthologs to 177 of the 289 human disease genes examined and provides the foundation for rapid analysis of some of the basic processes involved in human disease.

With the full genomic sequence of three major model organisms now available, much of our knowledge about the evolutionary basis of cellular and developmental processes will derive from comparisons between protein domains, intracellular networks, and cell-cell interactions in different phyla. In this paper, we begin a comparison of *D. melanogaster*, *C. elegans*, and *S. cerevisiae*. We first ask how many distinct protein families each genome encodes, how the genes encoding these protein families are distributed in each genome, and how many genes are shared among flies, worms, yeast, and mammals. Next we describe the composition and organization of protein domains within the proteomes of fly, worm, and yeast and examine the representation in each genome of a subset of genes that have been directly implicated as causative agents of human disease. Then we compare some fundamental cellular and developmental processes: the cell cycle, cell structure, cell adhesion, cell signaling, apoptosis, neuronal signaling, and the immune system. In each case, we present a summary of what we have learned from the sequence of the fly genome and how the components that carry out these processes differ in other organisms. We end by presenting some observations on what we have learned, the obvious questions that remain, and how knowledge of the sequence of the *Drosophila* genome will help us approach new areas of inquiry.

The "Core Proteome"

How many distinct protein families are encoded in the genomes of *D. melanogaster*, *C. elegans*, and *S. cerevisiae* (*1*), and how do these genomes compare with that of a simple prokaryote, *Haemophilus influenzae*? We carried out an "all-against-all" comparison of protein sequences encoded by each genome using algorithms that aim to differentiate paralogs—highly similar proteins that occur in the same genome—from proteins that are uniquely represented (Table 1). Counting each set of paralogs as a unit reveals the "core proteome": the number of distinct protein families in each organism. This operational definition does not include posttranslationally modifed forms of a protein or isoforms arising from alternate splicing.

In *Haemophilus*, there are 1709 protein coding sequences, 1247 of which have no sequence relatives within *Haemophilus* (*2*). There are 178 families that have two or more paralogs, yielding a core proteome of 1425. In yeast, there are 6241 predicted proteins and a core proteome of 4383 proteins. The fly and worm have 13,601 and 18,424 (*3*) predicted protein-coding genes, and their core proteomes consist of 8065 and 9453 proteins, respectively. It is remarkable that *Drosophila*, a complex metazoan, has a core proteome only twice the size of that of yeast. Furthermore, despite the large differences between fly and worm in terms of development and morphology, they use a core proteome of similar size.

Gene Duplications

Much of the genomes of flies and worms consists of duplicated genes; we next asked how these paralogs are arranged. The frequency of local gene duplications and the number of their constituent genes differ widely between fly and worm, although in both genomes most paralogs are dispersed. The fly genome contains half the number of local gene duplications relative to *C. elegans* (*4*), and these gene clusters are distributed randomly along the chromosome arms; in *C. elegans* there is a concentration of gene duplications in the recombinogenic segments of the autosomal arms (*1*). In both organisms, approximately 70% of duplicated gene pairs are on the same strand (306 out of 417 for *D. melanogaster* and 581 out of 826 for *C. elegans*). The largest cluster in the fly contains 17 genes that code for proteins of unknown function; the next largest clusters both consist of glutathione S-transferase genes, each with 10 members. In contrast, 11 of 33 of the largest clusters in *C. elegans* consist of genes coding for seven transmembrane domain receptors, most of which are thought to be involved in chemosensation. Other than these local tandem duplications, genes with similar functional assignment in the Gene Ontology (GO) classification (*5*) do not appear to be clustered in the genome.

We next compared the large duplicated gene families in fly, worm, and yeast without regard to genomic location. All of the known and predicted protein sequences of these three genomes were pooled, and each protein was compared to all others in the pool by means of the program BLASTP. Among the larger protein families that are found in worms and flies but not yeast are several that are associated with multicellular development, including homeobox proteins, cell adhesion molecules, and guanylate cyclases, as well as trypsinlike peptidases and esterases. Among the large families that are present only in flies are proteins involved in the immune response, such as lectins and peptidoglycan recognition proteins, transmembrane proteins of unknown function, and proteins that are probably fly-specific: cuticle proteins, peritrophic membrane proteins, and larval serum proteins.

Gene Similarities

What fraction of the proteins encoded by these three eukaryotes is shared? Comparative analysis of the predicted proteins encoded by these genomes suggests that nearly 30% of the fly genes have putative orthologs in the worm genome. We required that a protein show significant similarity over at least 80% of its length to a sequence in another species to be considered its ortholog (*6*). We know that this results in an underestimate, because the length requirement excludes known orthologs, such as homeodomain proteins, which have little similarity outside the homeodomain. The number of such fly-worm pairs does not decrease much as the similarity scores become more stringent (Table 2A), which strongly suggests that we have indeed identified orthologs, which may share molecular function. Nearly 20% of the fly proteins have a putative ortholog in both worm and yeast; these shared proteins probably perform functions common to all eukaryotic cells.

We also compared the proteins of fly, worm, and yeast to mammalian sequences. Most mammalian sequences are available as short expressed sequence tags (ESTs), so we dispensed with the requirement for similarity over 80% of the length of the proteins. Table 2B presents these data. Half of the fly protein sequences show similarity to mammalian proteins at a cutoff of $E < 10^{-10}$ (where E is expectation value), as compared to only 36% of worm proteins. This difference increases as the criteria become more stringent: 25% versus 15% at $E < 10^{-50}$ and 12% versus 7% at $E < 10^{-100}$. Because many of the comparisons are with short sequences, it is likely that many of these sequence similarities reflect conserved domains within proteins rather than orthology. However, it does suggest that the *Drosophila* proteome is more similar to mammalian proteomes than are those of worm or yeast.

Protein Domains and Families

Proteins are often mosaic, containing two or more different identifiable domains, and domains can occur in different combinations in different proteins. Thus, only a portion of a protein may be conserved among organisms. We therefore performed a comparative analysis of the protein domains composing the predicted proteomes from *D. melanogaster*, *C. elegans*, and *S. cerevisiae* using sequence similarity searches against the SWISS-PROT/TrEMBL nonredundant protein database (*7*), the BLOCKS database (*8*), and the InterPro database (*9*). The 200 most common fly protein families and domains are listed in Table 3, and the 10 most highly represented families in worm and yeast are shown in Table 4. InterPro analyses plus manual data inspection enabled us to assign 7419 fly proteins, 8356 worm proteins, and 3056 yeast proteins to either protein families or domain families. We found 1400 different protein families or domains in all: 1177 in the fly, 1133 in the worm, and 984 in yeast; 744 families or domains were common to all three organisms.

Many protein families exhibit great disparities in abundance, and only the C2H2-type zinc finger proteins and the eukaryotic

[1]Howard Hughes Medical Institute, [2]Department of Molecular and Cell Biology, Berkeley *Drosophila* Genome Project, University of California, Berkeley, CA 94720, USA. [3]Celera Genomics, Rockville, MD, 20850 USA. [4]GenetixXpress, 78 Pacific Road, Palm Beach, Sydney, Australia 2108. [5]Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, MA 02129 USA. [6]Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [7]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [8]Department of Genetics, Stanford University, Palo Alto, CA 94305, USA. [9]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [10]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. [11]Neurogenetics Unit, Laboratory of Neurochemistry, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA. [12]Berkeley *Drosophila* Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [13]Neomorphic, 2612 Eighth Street, Berkeley, CA 94710, USA. [14]School of Biology, University of Leeds, Leeds LS2 9JT, UK. [15]Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA. [16]Departments of Cellular and Molecular Medicine and Pharmacology, Howard Hughes Medical Institute, University of California–San Diego, La Jolla, CA 92093, USA. [17]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. [18]Howard Hughes Medical Institute, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [19]Genome Sequence Centre, BC Cancer Research Centre, 600 West 10th Avenue, Vancouver, BC, V52 4E6, Canada. [20]Molecular and Cell Biology Program, University of Maryland at Baltimore, Baltimore, MD 21201, USA. [21]Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette, France. [22]Center for Learning and Memory, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [23]Regulation of Cell Growth Laboratory, Division of Basic Sciences, National Cancer Institute–Frederick Cancer Research and Development Center, National Institutes of Health, Frederick, MD 21702, USA. [24]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. [25]Center for Neurobiology and Behavior, Columbia University, New York, NY 10032, USA. [26]Baylor College of Medicine Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

Table 1. Numbers of distinct gene families versus numbers of predicted genes and their duplicated copies in *H. influenzae*, *S. cerevisiae*, *C. elegans*, and *D. melanogaster*. Row one shows the total number of genes in each species. Row two shows the total number of all genes in each genome that appear to have arisen by gene duplication. Row three is the total number of distinct gene families for each genome. Each proteome was compared to itself using the same parameters as described in (*63*).

	H. influenzae	*S. cerevisiae*	*C. elegans*	*D. melanogaster*
Total no. of predicted genes	1709	6241	18424	13601
No. of genes duplicated	284	1858	8971	5536
Total no. of distinct families	1425	4383	9453	8065

protein kinases are among the top 10 protein families common to all three organisms. There are 352 zinc finger proteins of the C2H2 type in the fly but only 138 in the worm; whether this reflects greater regulatory complexity in the fly is not known. The protein kinases constitute approximately 2% of each proteome. Curation of the genomic data revealed that *Drosophila* has approximately 300 protein kinases and 85 protein phosphatases, around half of which had previously been identified. In contrast, there are approximately 500 kinases and 185 phosphatases in the worm; the difference is largely due to the worm-specific expansion of certain families such as the CK1, FER, and KIN-15 families. There are currently approximately 600 kinases and 130 phosphatases in humans, and it is expected that these figures will rise to 1100 and 300, respectively, when the sequence of the human genome is completed (*10*). Of the proteins uncovered in this analysis, over 70% exhibit sequence similarity outside the kinase or phosphatase domain to proteins in other species. In the kinase group, approximately 75% are serine/threonine kinases, and 25% are tyrosine or dual-specificity kinases. Over 90% of the newly discovered kinases are predicted to phosphorylate serine/threonine residues; this group includes the first atypical protein kinase C isoforms identified in *Drosophila*. In addition, we found counterparts of the mammalian kinases CSK, MLK2, ATM, and Peutz-Jeghers syndrome kinase, and additional members of the *Drosophila* GSK3B, casein kinase I, SNF1-like, and Pak/STE20-like kinase families. In the fly protein phosphatase group, approximately 42% are predicted to be serine/threonine phosphatases; 48% are tyrosine or dual-specificity phosphatases. Among the newly discovered phosphatases, 35% are serine/threonine phosphatases, most of which are related to the protein phosphatase 2C family, and 65% are tyrosine or dual-specificity phosphatases. The fly and worm both contain close relatives to many of the known mammalian lipid kinases and phosphatases; however, no SH2-containing inositol 5′ phosphatase SHIP is apparent. Finally, it has been found that the assembly of kinase signaling complexes in vertebrate cells is aided by the presence of scaffolding and adaptor molecules, many of which contain phosphoprotein binding domains; we found 85 such proteins in the fly, including counterparts to IRS, VAV, SHC, JIP, and MP1.

Two remarkable findings emerge from the peptidase data that may reflect different approaches to growth and development in flies, worms, and humans. The pattern and distribution of peptidase types are similar between the fly and the worm: there are approximately 450 peptidases in the fly and 260 in the worm. The difference is due almost entirely to the expansion or contraction of a single class of trypsin-like (S1) peptidases. *C. elegans* has seven of this class and yeast has one, but the fly has 199. Of these, 163 are small proteins of approximately 250 amino acids containing single trypsin domains; very few are mosaic proteins. The remainder have either multiple trypsin-like domains or long stretches of amino acids with no readily identifiable motif, usually at the NH_2-terminus. In humans, trypsin-like peptidases perform diverse functions in digestion, in the complement cascade, and in several other signaling pathways (*11*), and flies may have a similarly wide range of uses for these proteins. The extensively characterized members of this family, which include Snake, Easter, Nudel, and Gastrulation-defective, are all key members of a regulatory cascade that controls dorsoventral patterning in the fly (*12*). In addition, flies have only two members of the M10 class of peptidases, which include the matrix metalloproteases, collagenases, and gelatinases that are essential for tissue remodeling and repair in vertebrates.

The number of identifiable multidomain proteins is similar in the fly and the worm: 2130 and 2261, respectively. Yeast has only 672 (Table 5). Part of this difference is ac-

Table 2A. Similarity of sequences in predicted proteomes of *D. melanogaster*, *S. cerevisiae*, and *C. elegans*. To be scored as a similarity, each pairwise similarity was required to extend over more than 80% of the length of the query sequence at an E value less than that indicated. For example, in "Fly proteins in Fly-yeast," the column labeled $E < 10^{-10}$ shows the number and percentage of fly proteins that match yeast proteins at this E value or less and for which more than 80% of the length of the fly protein is aligned with the yeast protein. Each set of pairs was analyzed without consideration of the third proteome. The rows labeled "Fly-worm-yeast" report the composition of an independent clustering in which only groups containing a member from all three proteomes were counted. The numbers are slightly higher for the "Fly-worm-yeast" counts than for the "Fly-yeast" or "Worm-yeast" counts because of sequence bridging; that is, not all sequences within a group necessarily have a significant match to all other members of that group. See (*6*) for details.

	$E < 10^{-10}$		$E < 10^{-20}$		$E < 10^{-50}$		$E < 10^{-100}$	
	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)
Fly proteins in:								
Fly-yeast	2345	16.5	1877	13.2	1036	7.3	433	3.1
Fly-worm	4998	35.2	4212	29.7	2442	17.2	1106	7.8
Fly-worm-yeast	3303	23.3	2428	17.1	1113	7.8	435	3.1
Worm proteins in:								
Worm-yeast	2184	11.8	1768	9.5	933	5.0	374	2.0
Fly-worm	4795	25.8	4004	21.6	2403	12.9	1092	5.9
Fly-worm-yeast	3229	17.4	2439	13.1	1115	6.0	419	2.3
Yeast proteins in:								
Fly-yeast	1856	29.4	1567	24.8	891	14.1	376	6.0
Worm-yeast	1704	27.0	1425	22.6	802	12.7	335	5.3
Fly-worm-yeast	1833	29.1	1525	24.2	831	13.2	352	5.6

Table 2B. A comparison of *D. melanogaster*, *C. elegans*, and *S. cerevisiae* protein sequences to each other and to mammalian sequences (*64*). This table reports the number and percent of fly, worm, or yeast query sequences with similarities less than the indicated E value cutoffs. For example, in the "Fly vs. Yeast" comparison, 3986 or 28.1% of fly proteins have a similarity with a yeast protein with an E value less than 1×10^{-10}. EST E values are not directly comparable to protein E values, because the resulting alignments are shorter.

	No similarity $E > 10^{-4}$		$E < 10^{-10}$		$E < 10^{-20}$		$E < 10^{-50}$		$E < 10^{-100}$	
	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)
Fly vs.										
Yeast	8177	57.6	3986	28.1	2677	18.9	1266	8.9	504	3.6
Worm	5110	36.0	6743	47.5	5180	36.5	2832	19.9	1197	8.4
Mammalian	5833	41.1	7032	49.5	5837	41.1	3580	25.2	1772	12.5
Mammalian ESTs	5386	37.9	7329	51.6	5352	37.7	1775	12.5	110	0.8
Worm vs.										
Yeast	12541	68.0	3582	19.4	2378	12.9	1106	6.0	401	2.2
Fly	8603	46.7	7138	38.8	5428	29.5	2880	15.6	1229	6.7
Mammalian	10152	55.1	6550	35.6	4999	27.1	2782	15.1	1211	6.6
Mammalian ESTs	10354	56.2	6005	32.6	4000	21.7	1170	6.4	68	0.4
Yeast vs.										
Fly	2614	41.9	2564	41.0	1910	30.6	1021	16.4	408	6.5
Worm	2762	44.2	2358	37.8	1730	27.7	882	14.1	348	5.6
Mammalian	3230	51.7	2340	37.5	1802	28.9	992	15.9	429	6.9
Mammalian ESTs	3106	49.7	2319	37.1	1553	24.9	503	8.1	18	0.3

counted for by proteins with extracellular domains involved in cell-cell and cell-substrate contacts (*13*), such as the immunoglobulin domain–containing proteins, which are more abundant in flies than in worms (153 versus 70) and are nonexistent in yeast. Two other common extracellular domains occur in similar numbers in fly and worm: EGF (110 versus 109, respectively) and fibronectin type III (46 versus 43) but are rare or absent in yeast. Extracellular regions of proteins often contain a variety of repeated domains (*14*), and so these proteins may account for our finding that flies have a larger number of proteins with multiple InterPro domains than either worms or yeast (2107 versus 1747 and 525, respectively) (Table 6). Some multidomain proteins of the fly are particularly heterogeneous: Two low-density lipoprotein receptor–related proteins have 75 InterPro domains each. Another protein of unknown function has 62 InterPro domains; the most heterogeneous worm and yeast proteins [SWISS-PROT/TrEMBL accession numbers (AC), Q04833 and P32768, respectively] have 61 and 18 InterPro domains, respectively. There can be extensive repetition of the same domain within a protein; for example, an immunoglobulin-like domain is repeated 52 times within one protein of unknown function in the fly. The large worm protein UNC-89 contains 48 immunoglobulin-like domains (SWISS-PROT/TrEMBL AC, Q17362). In contrast, the largest number of repeats in yeast, of a C2H2-type zinc finger domain, occurs nine times in the transcription factor TFIIIA (SWISS-PROT/TrEMBL AC, P39933).

The heterotrimeric GTP-binding protein (G protein)–coupled receptors (GPCRs) are a large protein family in flies, worms, and vertebrates whose members are involved in synaptic function, hormonal physiology, and the regulation of morphological movements during gastrulation and germ band extension (*15*). There are predicted to be at least 700 GPCRs in the human genome (*16*) and roughly 1100 GPCRs in *C. elegans* (*17*). We found approximately 160 GPCR genes in the *Drosophila* genome, 57 of which appear to be olfactory receptors. *Drosophila*, *C. elegans*, and vertebrates each have diverse families of odorant receptors that, although recognizable as GPCRs, are unrelated by sequence and therefore apparently evolved independently. The number of odorant receptors in vertebrates ranges from around 100 in zebrafish and catfish to approximately 1000 in the mouse; *C. elegans* also has approximately 1000. In the fly, as in zebrafish and mouse, there is a correlation between the number of odorant receptors and the number of discrete synaptic structures called glomeruli in the olfactory processing centers of the brain (*16*, *18*). In the mouse, each glomerulus is dedicated to receiving axonal input from neurons expressing a particular odorant receptor (*16*). Therefore, the correlation between number of odorant receptors and number of glomeruli may reflect a conservation in the organizational logic of odor recognition in insect and vertebrate brains. Although the fly odorant receptors are extremely diverse, there are a number of subfamilies whose members share 50 to 65% sequence identity. The distribution of odorant receptor genes is different among these organisms as well. Unlike *C. elegans* or vertebrate odorant receptors, which are in large linked arrays, the fly odorant receptor genes are distributed as single genes or in arrays of two or three. Vertebrate receptors are encoded by intronless genes, but both fly and worm receptor genes have multiple introns. These distinctions suggest that in addition to differences in the sequences of the odorant receptors of the different organisms, the processes generating the families of receptors may have differed among the lineages that gave rise to flies, worms, and vertebrates.

The data suggest conservation of hormone receptors between flies and vertebrates; nevertheless, there is a greater diversity of hormone receptors in both *C. elegans* and vertebrates than in *Drosophila*. Insects are subject to complex hormonal regulation, but no apparent homologs of vertebrate neuropeptide and hormone precursors were identified. However, many receptors with sequence similarity to vertebrate receptors for neurokinin, growth hormone secretagogue, leutotropin (follicle-stimulating hormone and luteinizing hormone), thyroid-stimulating hormone, galanin/allatostatin, somatostatin, and vasopressin were identified. Other GPCRs include a seventh *Drosophila* rhodopsin and homologs of adenosine, metabotropic glutamate, γ-aminobutyric acid (GABA), octopamine, serotonin, dopamine, and muscarinic acetylcholine receptors. In addition, there are GPCRs that are unique to *Drosophila*, others with sequence similarity to *C. elegans* and human orphan receptors, and an insect diuretic hormone receptor that is closely related to vertebrate corticotropin-releasing factor receptor. Finally, we found several atypical seven-transmembrane domain receptors, including 10 Methuselah (MTH)–like proteins and four Frizzled (FZ)–like proteins. A mutation in *mth* increases the fly's life-span and its resistance to various stresses (*19*); the FZ-like proteins probably serve as receptors for different members of the Wingless/Wnt family of ligands.

Human Disease Genes

Studies in model organisms have provided important insights into our understanding of genes and pathways that are involved in a variety of human diseases. In order to estimate the extent to which different types of human disease genes are found in flies, worms, and yeast, we compiled a set of 289 genes that are mutated, altered, amplified, or deleted in a diverse set of human diseases and searched for similar genes in *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, as described in the legend to Fig. 1. Of these 289 human genes, 177 (61%) appear to have an ortholog in *Drosophila* (Fig. 1). Only proteins with similar domain structures were considered to be orthologs; this judgment was made by human inspection of the InterPro domain composition of the fly and human proteins. The importance of human inspection, as well as consideration of published information, is underscored by the fact that some sequences with extremely high similarity scores to proteins encoded by fly genes, such as LCK and Myotonic Dystrophy 1, were judged not to be orthologous, but others with relatively low scores, such as p53 and Rb1, were considered to be orthologs. We attempted this additional level of analysis only for the fly proteins, as the lower overall level of similarity of worm and yeast proteins made these subjective judgments even more difficult. Some of the human disease genes that are absent in *Drosophila* reflect clear differences in physiology between the two organisms. For instance, none of the hemoglobins, which are mutated in thalassemias, have orthologs in *Drosophila*. In flies, oxygen is delivered directly to tissues via the tracheal system rather than by circulating erythrocytes. Similarly, several genes required for normal rearrangement of the immunoglobulin genes do not have *Drosophila* orthologs.

Of the cancer genes surveyed, 68% appear to have *Drosophila* orthologs. In addition to previously described proteins, these searches identified clear protein orthologs for menin (MEN; multiple endocrine neoplasia type 1), Peutz-Jeghers disease (STK11), ataxia telangiectasia (ATM), multiple exostosis type 2 (EXT2), a second bCL2 family member, a second retinoblastoma family member, and a p53-like protein. Despite its relatively low sequence similarity to the human genes, the *Drosophila* gene encoding p53 was considered an ortholog because it shows a conserved organization of functional domains, and its DNA binding domain includes many of the same amino acids that appear to be hot spots for mutations in human cancer. Comparison of the fly p53-like protein with the human p53, p63, and p73 proteins suggests that it may represent a progenitor of this entire family. In mammalian cells, levels of p53 protein are tightly regulated in vivo by its interaction with the Mdm2 protein, which in turn binds to p19ARF (*20*). This mode of regulation, which modulates the activity of p53 but probably not of p63 or p73 (*21*), may not apply to the *Drosophila* protein, because we have not been able to identify orthologs of

Table 3. Number of proteins in *D. melanogaster* (F), *C. elegans* (W), and *S. cerevisiae* (Y) containing the 200 most frequently occurring protein domains in *D. melanogaster*. Domain identifiers are from InterPro (9), a new database that has begun to integrate the independent databases of localized protein sequence patterns into a single resource. The beta release used includes PROSITE, PRINTS, and PFAM. InterPro considers a signature to be true if its score is above a threshold specified for that signature by the individual database. Results of the InterPro analysis may differ from results obtained based on human curation of protein families, due to the limitations of large-scale automatic classifications. In some instances, different InterPro domains correspond to different features of proteins within the same family; for example, IPR001650 and IPR001410 (26 and 42 in the table). See (62) for live links to the InterPro database.

	Acc. No.	F	W	Y	Interpro Domain Name
1.	IPR000694	579	398	40	Proline-rich region
2.	IPR000822	352	138	47	Zinc finger, C2H2 type
3.	IPR000719	249	388	119	Eukaryotic protein kinase
4.	IPR001254	199	13	1	Serine proteases, trypsin family
5.	IPR001314	178	5	0	Chymotrypsin serine protease family (S1)
6.	IPR001680	167	95	90	G-protein beta WD-40 repeats
7.	IPR000504	160	92	55	RNA-binding region RNP-1 (RNA recognition motif)
8.	IPR000495	153	70	0	Immunoglobulins & major histocompatibility complex proteins
9.	IPR000345	145	17	7	Cytochrome c family heme-binding site
10.	IPR000379	140	112	38	Esterase/lipase/thioesterase
11.	IPR002290	138	171	110	Serine/Threonine protein kinases active-site
12.	IPR002048	130	79	16	EF-hand family
13.	IPR001356	113	88	10	Homeobox domain
14.	IPR000561	110	109	0	EGF-like domain
15.	IPR001611	108	48	7	Leucine-rich repeat
16.	IPR001841	105	113	35	Zinc finger, C3HC4 type (RING finger)
17.	IPR002356	100	335	0	G-protein coupled receptors, rhodopsin family
18.	IPR001066	97	54	46	Sugar transporter
19.	IPR001128	94	73	3	Cytochrome P450 enzyme
20.	IPR002110	90	77	19	Ankyrin-repeat
21.	IPR000618	87	0	0	Insect cuticle protein
22.	IPR001245	87	63	0	Tyrosine kinase catalytic domain
23.	IPR001440	82	46	34	TPR repeat
24.	IPR000130	79	19	8	Neutral zinc metallopeptidases, zinc-binding region
25.	IPR002380	78	41	22	Transforming protein P21 RAS
26.	IPR001650	76	66	75	DNA/RNA helicase domain (DEAD/DEAH box)
27.	IPR001617	72	56	32	ABC transporters family
28.	IPR001849	71	67	27	PH domain
29.	IPR001478	69	60	2	PDZ domain (also known as DHR or GLGF)
30.	IPR001488	69	8	5	Myc-type, helix-loop-helix dimerization domain signature
31.	IPR001051	67	61	38	ATP-binding transport protein, 2nd P-loop motif
32.	IPR001993	67	43	35	Mitochondrial energy transfer proteins
33.	IPR000734	66	9	4	Lipase
34.	IPR000210	64	103	1	Btb/ttk domain
35.	IPR000575	63	54	36	ATP/GTP-binding site motif A (P-loop)
36.	IPR001452	63	55	25	Src homology 3 (SH3) domain
37.	IPR001092	61	38	8	Helix-loop-helix DNA-binding domain
38.	IPR002198	61	63	13	Short-chain dehydrogenase/reductase (SDR) superfamily
39.	IPR002106	58	14	17	Aminoacyl-transfer RNA synthetases class-II
40.	IPR001806	51	46	23	Ras family
41.	IPR002347	50	22	1	Glucose/ribitol dehydrogenase family
42.	IPR001410	46	43	48	DEAD/DEAH box helicase
43.	IPR001777	46	43	2	Fibronectin type III domain
44.	IPR000169	43	22	1	Eukaryotic thiol (cysteine) proteases active sites
45.	IPR000521	42	44	6	Glutathione S-transferase
46.	IPR001622	42	91	1	Potassium channel
47.	IPR002557	42	6	0	Chitin binding domain
48.	IPR000051	40	38	21	SAM (and some other nucleotide) binding motif
49.	IPR002172	40	32	0	Low density lipoprotein (LDL)-receptor class A (LDLRA) domain
50.	IPR000063	38	32	12	Thioredoxin family
51.	IPR001623	38	29	22	DnaJ domain
52.	IPR002018	38	44	0	Carboxylesterases type-B
53.	IPR001304	37	165	0	C-type lectin domain
54.	IPR000387	36	83	12	Tyrosine specific protein phosphatase
55.	IPR000215	35	9	0	Serpins
56.	IPR001005	35	16	19	Myb DNA binding domain
57.	IPR001412	35	15	14	Aminoacyl-transfer RNA synthetases class-I
58.	IPR001939	35	27	29	AAA-protein (ATPases associated with various cellular activities)
59.	IPR001965	35	22	16	PHD-finger
60.	IPR000008	34	34	9	Protein kinase C2 domain
61.	IPR000608	34	18	16	Ubiquitin-conjugating enzymes
62.	IPR001781	34	33	4	LIM domain
63.	IPR000980	33	43	1	Src homology 2 (SH2) domain
64.	IPR002213	33	59	0	UDP-glucoronosyl & UDP-glucosyl transferases
65.	IPR000301	32	19	0	Transmembrane 4 family
66.	IPR000934	31	56	21	Serine/threonine specific protein phosphatase family
67.	IPR001251	31	16	6	CRAL/TRIO domain
68.	IPR001881	31	34	0	Calcium-binding EGF-like domain
69.	IPR002173	31	4	2	PfkB family of carbohydrate kinases
70.	IPR000194	30	5	22	ATP synthase alpha & beta subunits
71.	IPR000217	29	22	4	Tubulin family
72.	IPR000873	29	23	11	AMP-binding domain
73.	IPR000073	28	17	16	Alpha/beta hydrolase fold
74.	IPR000152	28	28	0	Aspartic acid & asparagine hydroxylation site
75.	IPR000408	28	6	3	Regulator of chromosome condensation (RCC1)
76.	IPR000834	28	9	1	Zinc carboxypeptidases, carboxypeptidase A metalloprotease (M14) family
77.	IPR001715	28	22	3	Calponin homology (CH) domain
78.	IPR002086	28	13	13	Aldehyde dehydrogenase family
79.	IPR002219	28	36	1	Phorbol esters/diacylglycerol binding domain
80.	IPR000483	27	7	0	Leucine rich repeat C-terminal domain
81.	IPR000886	27	8	11	Endoplasmic reticulum targeting sequence
82.	IPR001175	27	81	0	Neurotransmitter-gated ion-channel
83.	IPR000219	26	17	5	Dbl domain (dbl/cdc24 rhoGRF family)
84.	IPR000626	26	27	9	Ubiquitin domain
85.	IPR000629	26	22	20	ATP-dependent helicase, DEAD-box subfamily
86.	IPR000859	26	55	0	CUB domain
87.	IPR000958	26	21	6	KH domain
88.	IPR001752	26	22	6	Kinesin motor domain
89.	IPR002067	26	11	6	Mitochondrial carrier protein
90.	IPR000205	25	22	10	NAD binding site
91.	IPR000299	25	13	0	Band 4.1 family
92.	IPR000449	25	10	8	Ubiquitin-associated domain
93.	IPR000910	25	15	8	HMG1/2 (high mobility group) box
94.	IPR001054	25	32	1	Guanylate cyclase
95.	IPR001202	25	17	5	WW/rsp5/WWP domain
96.	IPR000595	24	19	2	Cyclic nucleotide-binding domain
97.	IPR000832	24	10	0	G-protein coupled receptors family 2 (secretin-like)
98.	IPR001140	24	30	10	ABC transporter transmembrane region
99.	IPR001214	24	27	6	SET-domain of transcriptional regulators (TRX, EZ, ASH1 etc)
100.	IPR001871	24	18	15	bZIP (Basic-leucine zipper) transcription factor family
101.	IPR002049	23	16	0	Laminin-type EGF-like (LE) domain
102.	IPR002111	23	21	2	Cation channels, 6TM region (transient receptor potential subtype)
103.	IPR000048	22	16	2	IQ calmodulin-binding domain
104.	IPR001353	22	12	14	Multispecific proteases of the proteasome
105.	IPR001810	22	215	11	F-box domain
106.	IPR002223	22	34	0	Pancreatic trypsin inhibitor (Kunitz) family
107.	IPR000718	21	29	0	Neprilysin metalloprotease (M13) family
108.	IPR000964	21	15	3	Sterile-alpha module (SAM) domain
109.	IPR001311	21	13	0	Solute binding protein/glutamate receptor domain
110.	IPR001394	21	24	18	Ubiquitin carboxyl-terminal hydrolases family 2
111.	IPR001594	21	13	6	DHHC-type Zn-finger
112.	IPR001628	21	224	0	C4-type steroid receptor zinc finger
113.	IPR002017	21	19	3	Spectrin repeat
114.	IPR002113	21	6	4	Adenine nucleotide translocator 1
115.	IPR002126	21	15	0	Cadherin domain
116.	IPR000195	20	17	12	RabGAP/TBC domain
117.	IPR000198	20	19	10	RhoGAP domain
118.	IPR000795	20	17	15	GTP-binding elongation factor
119.	IPR001930	20	11	4	Membrane alanyl dipeptidase, family M1
120.	IPR002422	20	14	7	Permeases for amino acids & related compounds, family II
121.	IPR000166	19	33	16	Histone-fold/TFIID-TAF/NF-Y domain
122.	IPR000690	19	8	7	RNA-binding protein C2H2 Zn-finger domain
123.	IPR001766	19	19	4	Fork head domain
124.	IPR002130	19	17	8	Cyclophilin-type peptidyl-prolyl cis-trans isomerase
125.	IPR002293	19	16	25	Permeases for amino acids & related compounds, family I
126.	IPR000175	18	12	0	Sodium:neurotransmitter symporter family
127.	IPR000330	18	20	17	SNF2 & others N-terminal domain

Table 3 (continued).

	Acc. No.	F	W	Y	Interpro Domain Name
128.	IPR000742	18	9	0	EGF-like domain, subtype 2
129.	IPR000961	18	24	10	Protein kinase C terminal domain
130.	IPR001173	18	17	4	Glycosyl transferase, family 2
131.	IPR000242	17	76	3	Tyrosine specific protein phosphatases
132.	IPR000467	17	11	4	D111 domain
133.	IPR000636	17	22	1	Cation channels, 6TM region (non-ligand gated)
134.	IPR000717	17	13	8	Domain in components of the proteasome, COP9-complex & eIF3 (PCI)
135.	IPR000953	17	15	2	Chromo domain
136.	IPR001071	17	0	0	Alpha-tocopherol transport protein
137.	IPR001163	17	11	16	Small nuclear ribonucleoprotein (Sm protein)
138.	IPR001327	17	4	4	FAD-dependent pyridine nucleotide reductase
139.	IPR001395	17	11	6	Aldo/keto reductase family
140.	IPR001734	17	3	1	Sodium:solute symporter family
141.	IPR001757	17	22	17	E1-E2 ATPases phosphorylation site
142.	IPR001791	17	16	0	Laminin-G domain
143.	IPR001873	17	22	0	Amiloride-sensitive sodium channel
144.	IPR001969	17	8	42	Eukaryotic & viral aspartyl proteases active site
145.	IPR000087	16	166	0	Collagen triple helix repeat
146.	IPR000253	16	6	16	Forkhead-associated (FHA) domain
147.	IPR000536	16	88	0	Ligand-binding domain of nuclear hormone receptor
148.	IPR001320	16	10	0	Ligand-gated ion channel
149.	IPR001487	16	13	10	Bromodomain
150.	IPR002027	16	11	24	Amino acid permease
151.	IPR002046	16	1	1	SAR1 GTP-binding protein family
152.	IPR000014	15	8	1	Generalized PAS domain
153.	IPR000172	15	1	0	GMC oxidoreductases
154.	IPR000251	15	12	7	ADP-ribosylation factors family
155.	IPR000569	15	5	5	HECT-domain (Ubiquitin-transferase)
156.	IPR000772	15	12	0	Lectin domain of ricin b-chain, 3 copies
157.	IPR001223	15	34	1	Glycosyl hydrolases family 18
158.	IPR001609	15	20	5	Myosin head (motor domain)
159.	IPR001828	15	19	0	Receptor family ligand binding region
160.	IPR002129	15	7	1	Pyridoxal-dependent decarboxylase family
161.	IPR002465	15	1	0	Growth factor & cytokine receptor family signature 2
162.	IPR000159	14	11	2	Ras-associated (RalGDS/AF-6) domain
163.	IPR000225	14	6	2	Armadillo/plakoglobin ARM repeat
164.	IPR000279	14	10	8	Actin
165.	IPR000566	14	6	0	Lipocalin & cytosolic fatty-acid binding protein
166.	IPR000577	14	3	2	Carbohydrate kinase, FGGY family
167.	IPR000746	14	0	0	Pheromone/general odorant binding protein, PBP/GOBP family
168.	IPR000884	14	27	0	Thrombospondin type I domain
169.	IPR001100	14	5	3	Pyridine nucleotide-disulfide oxidoreductase, class I
170.	IPR001159	14	9	2	Double-stranded RNA binding (DsRBD) domain
171.	IPR001199	14	8	5	Cytochrome B5
172.	IPR001357	14	21	11	BRCT domain
173.	IPR001589	14	8	1	Actinin-type actin-binding domain
174.	IPR001753	14	10	3	Enoyl-CoA hydratase/isomerase
175.	IPR001878	14	24	9	Zn-finger CCHC type
176.	IPR001952	14	0	1	Alkaline phosphatase family
177.	IPR002216	14	17	1	Ion transport protein
178.	IPR002464	14	9	8	DEAH-box subfamily ATP-dependent helicase
179.	IPR000107	13	8	3	SPRY domain
180.	IPR000425	13	8	6	MIP family
181.	IPR000508	13	2	3	Signal peptidase
182.	IPR000727	13	14	15	t-SNARE coiled-coil domain
183.	IPR000901	13	6	7	Carbamoyl-phosphate synthase
184.	IPR001461	13	16	7	Pepsin (A1) aspartic protease family
185.	IPR001506	13	36	0	Astacin (Peptidase family M12A) family
186.	IPR001523	13	11	0	'Paired box' domain
187.	IPR001827	13	2	0	'Homeobox' antennapedia-type protein
188.	IPR001876	13	7	1	Zn-finger in ranbp & others
189.	IPR002423	13	8	9	TCP-1 (Tailless complex polypeptide)/cpn60 chaperonin family
190.	IPR002893	13	8	1	MYND finger
191.	IPR000461	12	4	8	Alpha amylase
192.	IPR000798	12	4	0	Ezrin/radixin/moesin family
193.	IPR001023	12	13	14	Heat shock protein hsp70
194.	IPR001508	12	1	0	NMDA receptor
195.	IPR001683	12	9	15	PX (Bem1/NCF1/PI3K) domain
196.	IPR001917	12	6	4	Aminotransferases class-II
197.	IPR001932	12	9	8	Protein phosphatase 2C
198.	IPR000050	11	8	0	Phosphotyrosine interaction domain (PID)
199.	IPR000182	11	9	9	cetyltransferase (GNAT) family
200.	IPR000243	11	2	7	Proteasome B-type subunit

either Mdm2 or p19ARF in *Drosophila*. Interestingly, likely orthologs of the breast cancer susceptibility genes *BRCA1* and *BRCA2* were not found in *Drosophila*. In most instances, cancer genes that have a *Drosophila* ortholog also have an ortholog in *C. elegans*, although the extent of sequence similarity to the worm gene is lower. In a minority of instances, a *C. elegans* ortholog was clearly absent. Cancer genes with orthologs in *Drosophila* and apparently not in *C. elegans* include *p53* and *neurofibromatosis type 1* (*22*), the two genes implicated in tuberous sclerosis (*TSC1* and *TSC2*) (*23*), and *MEN*. The two TSC gene products are thought to bind to each other and may function in a pathway that is conserved between humans and *Drosophila* but is absent in *C. elegans* and *S. cerevisiae*. However, the limitations of this type of analysis are clearly illustrated by our inability to find a *bCL2* ortholog in *C. elegans* using these search parameters. The *C. elegans ced-9* gene has been shown to function as a *bCL2* homolog, and its protein is 23% identical to the human protein over its entire length (*24*).

Numerous orthologs of neurological genes are also found in the *Drosophila* genome. Some, such as *Notch* (CADASIL syndrome), the *beta amyloid protein precursor-like* gene, and *Presenilin* (Alzheimer's disease), were already known from previous studies in the fly. The genome sequencing effort has uncovered several additional genes that are likely to be orthologs of human neurological genes, such as *tau* (frontotemporal dementia with Parkinsonism), the Best macular dystrophy gene, *neuroserpin* (familial encephalopathy), genes for limb girdle muscular dystrophy types 2A and 2B, the Friedreich ataxia gene, the gene for Miller-Dieker lissencephaly, *parkin* (juvenile Parkinson's disease), and the Tay-Sachs and Stargardt's disease genes. Several genes implicated in expanded polyglutamine repeat diseases, including Huntington's and spinal cerebellar ataxia 2 (*SCA2*), are found in the fruit fly. Most human neurological disease genes surveyed were also detected in *C. elegans*, and some were even found in yeast, although a few examples are apparently present only in *Drosophila*, such as the *Parkin* and *SCA2* orthologs.

Among genes implicated in endocrine diseases, those functioning in the insulin pathway are mostly conserved. In contrast, members of pathways involving growth hormone, mineralocorticoids, thyroid hormone, and the proteins that regulate body mass in vertebrates, such as those encoding leptin, do not appear to have *Drosophila* orthologs. Surprisingly, a protein that shows significant sequence similarity to the luteinizing hormone receptor is present in *Drosophila* (*25*). The physiological ligand for this receptor is not known. A number of genes that have been implicated in human renal disorders have orthologs in *Drosophila*, despite the differences between human kidneys and insect Malpighian tubules. In many instances, these gene products are involved in fluid and electrolyte transport across epithelia. Not surprisingly, most disease genes that function in intracellular metabolic pathways appear to have *Drosophila* orthologs.

Developmental and Cellular Processes

Developmental strategies in various phyla are overtly very different, from the fixed cell lineage of *C. elegans* to the syncytial embryogenic development of the fly, to early embryogenesis in amphibians and mammals. A number of major processes—cell division, cell shape, signaling pathways, cell-cell and

cell-substrate adhesion, and apoptosis—determine the developmental outcomes of these very different embryos. Although there are many more, such as the processes that determine embryonic gradients, cell polarities, and cell movement, here we examine the first five, beginning with cell cycle components, and examine what new insights have been gained from the genomic data that affect our knowledge of the evolution of developmental processes. We then discuss the processes of neuronal signaling and innate immunity.

Cell cycle. Despite conservation of the mechanisms regulating cell cycle progression, many of the functions governing this progression are encoded by gene families whose individual members are not conserved between vertebrates and yeast. For example, the cyclins of *S. cerevisiae* can be divided into a G_1 class (Cln1, Cln2, and Cln3) and an S/G_2 class (Clb1 through Clb6); it is not possible to identify orthologs of individual vertebrate cyclins. Consequently, analysis of the roles of particular vertebrate cell cycle genes benefits from a genetic model in which parallels are more evident. Analysis of the *Drosophila* genome sequence supports and extends previous suggestions of strong parallels between fly and human cell cycle regulators. Orthologs of vertebrate cell cycle cyclins—cyclin A (CycA), CycB, CycB3, CycE, and CycD—have been identified in *Drosophila*, as have orthologs of cyclins that appear to have roles in transcription: CycC, CycH, CycK, and CycT. Apparent orthologs of these cyclins can be also be found in *C. elegans*; however, the level of similarity to the vertebrate members is invariably substantially less. Indeed, BLAST comparisons suggest that vertebrate and *Drosophila* CycA and CycB share more sequence similarity with yeast than with proposed *C. elegans* orthologs. Examination of other cell cycle regulators confirms that quite precise comparisons can be made between vertebrates and flies; parallels with yeast are looser. For example, like vertebrates, *Drosophila* uses several different cyclin-dependent kinases (Cdks) to regulate different aspects of the cell cycle; *S. cerevisiae* and *Schizosaccharomyces pombe* use only one. Cloning efforts and the genome sequence revealed *Drosophila* orthologs of vertebrate Cdk1 (*cdc2*) and Cdk2 (*cdc2c*), as well as a single *Drosophila* Cdk (*Cdk4/6*) with close similarity to both Cdk4 and Cdk6. As in vertebrates, *Drosophila* has two distinct kinases that add inhibitory phosphate to Cdk1, the previously identified Wee, and a recently recognized homolog of Myt1, which was initially identified as a membrane-associated inhibitory kinase in *Xenopus* (*26*). *C. elegans* also has two homologs of these kinases (Wee1.1 and Wee1.3); however, similarity scores do not place these into distinct Wee1 and Myt1 subtypes. Each of these genes appears to be present in a single copy, a factor that simplifies genetic interpretations.

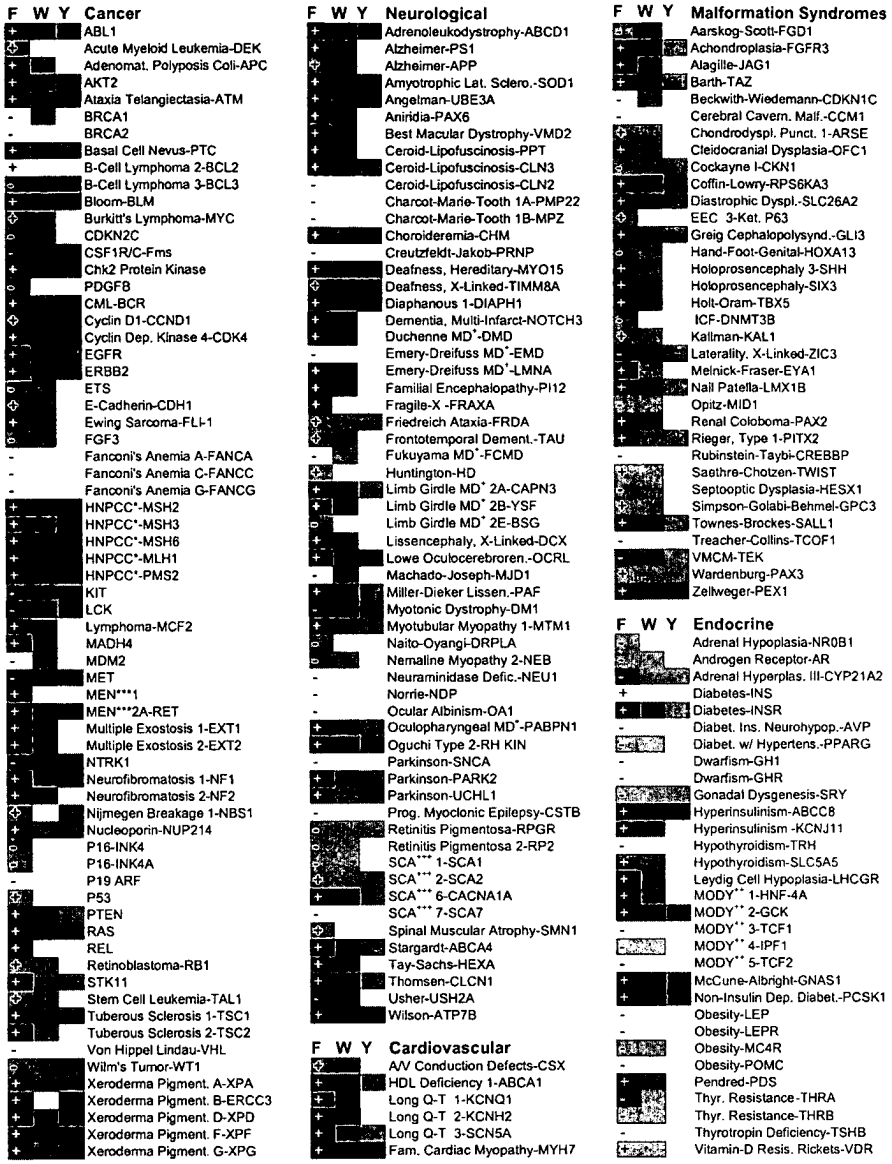


Fig. 1.

The retinoblastoma gene product pRb is a crucial cell cycle regulator in mammals and is thought to modulate S-phase entry via its interactions with the transcriptional regulator E2F and its dimerization partner (DP). This important mode of regulation is not found in yeast, but many components of the Rb pathway have been identified and studied in *Drosophila* (*27*). The sequencing effort uncovered a second *Rb*-related gene in *Drosophila* and confirmed the existence of only two E2F family members and a single DP ortholog. *C. elegans* also has an *Rb*-related gene, isolated in a genetic screen for mutations affecting cell fate decisions (*28*), but it has not been shown to play a direct role in cell cycle regulation. Also evident from the sequence are eight *skp*-like genes and six *cullin*-related genes. The Skp and Cullin proteins function in a complex that mediates the degradation of specific target proteins during crucial cell cycle transitions. Further exploration of the genome sequence should define orthologs to most vertebrate cell cycle genes and lead to genetic tests of their regulation and function.

Cytoskeleton. A large number of proteins link events at the cell surface with cytoskeletal networks and intracellular messengers (*13*). We found approximately 230 genes (approximately 2% of the predicted genes) that encode cytoskeletal structural or motor proteins; these represent most major families found in other invertebrates and vertebrates (*29*). The fraction of the *Drosophila* genome devoted to cytoskeletal functions appears to

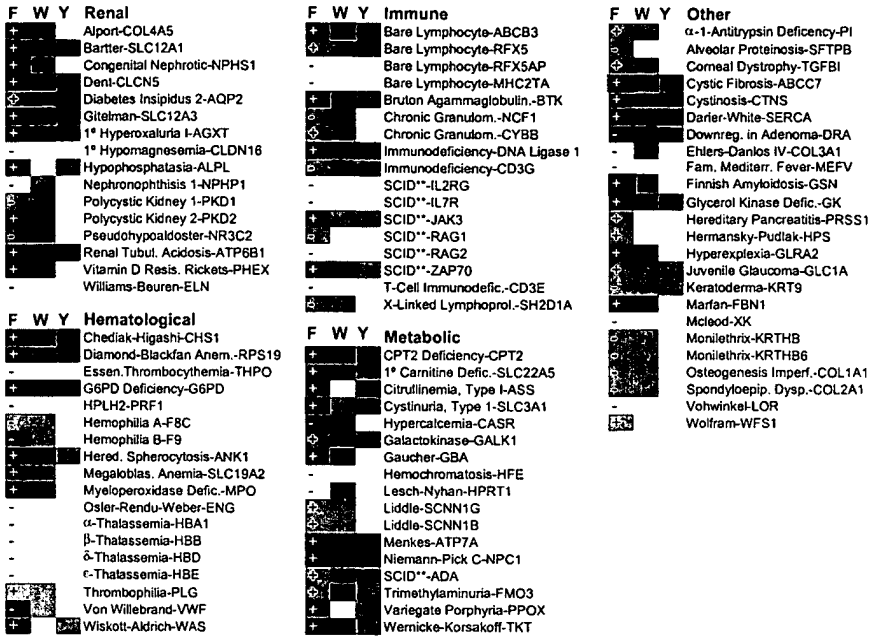


Fig. 1 (continued). Fly (F), worm (W), and yeast (Y) genes showing similarity to human disease genes. This collection of human disease genes was selected to represent a cross section of human pathophysiology and is not comprehensive. The selection criteria require that the gene is actually mutated, altered, amplified, or deleted in a human disease, as opposed to having a function deduced from experiments on model organisms or in cell culture. Due to redundancy in gene and protein sequence databases, a single reference sequence for each gene had to be chosen. Most reference sequences represent the longest mRNA of several alternatives in GenBank. Authoritative sources in the literature and electronic databases [Online Mendelian Inheritance in Man (OMIM)] were also consulted. In all, 289 protein sequences met these criteria. These were used as queries to search a database consisting of the sum total of gene products (38,860) found in the complete genomes of fly, worm, and yeast. 12,953 was used as the effective database size (the z parameter in BLAST). BLASTP searches were conducted as described for full genome searches, except for the z parameter. To control for potential frameshift errors in the *Drosophila* genome sequence, searches against a six-frame translation of the entire genome (using TBLASTN) were also conducted with the disease gene sequences using the z parameter above. Only two cases in which matches to genomic sequence were better than to the predicted protein were found, and these were manually corrected to reflect the better TBLASTN scores in the table. Results are scaled according to various levels of statistical significance, reflecting a level of confidence in either evolutionary homology or functional similarity. White boxes represent BLAST E values $>1 \times 10^{-6}$, indicating no or weak similarity; light blue boxes represent E values in the range of 1×10^{-6} to 1×10^{-40}; purple boxes represent E values in the range of 1×10^{-40} to 1×10^{-100}; and dark blue boxes represent E values $<1 \times 10^{-100}$, indicating the highest degree of sequence conservation. Actual E values can be found in the Web supplement to this figure (*62*), where links to OMIM and GenBank may also be found. A plus sign indicates our best estimate that the corresponding *Drosophila* gene product is the functional equivalent of the human protein, based on degree of sequence similarity, InterPro domain composition, and supporting biological evidence, when available. A minus sign indicates that we were unable to identify a likely functional equivalent of the human protein.

be somewhat smaller than that found in *C. elegans* (5%) (*30*); whether this reflects a true biological difference or a difference in classification criteria remains to be discovered. Of the *Drosophila* cytoskeletal genes, 90 encode proteins belonging to the kinesin, dynein, or myosin motor superfamilies, or accessory or regulatory proteins known to interact with the motor protein subunits. Approximately 80 genes encode actin-binding proteins, including proteins belonging to the spectrin/α-actinin/dystrophin superfamily of membrane cytoskeletal and actin–cross-linking proteins. Twenty genes encode proteins that are likely to bind microtubules, based on their similarity to microtubule-binding proteins found in other organisms. Fourteen genes encode members of the actin superfamily, 12 encode members of the tubulin superfamily, and 5 encode septins. Overall, the representation of predicted cytoskeletal protein types and families is similar to what has been found for *C. elegans*, although *Drosophila* has many more dyneins, probably because *C. elegans* lacks motile cilia and flagella.

Among this collection of cytoskeletal genes are several interesting and in some cases long-sought genes. One gene encodes a protein with striking homology to proteins of the tau/MAP2/MAP4 family that share a characteristic repeated microtubule-binding domain. Two encode new tubulins; one appears most closely related to α-tubulin, and the other appears most closely related to β-tubulin, both with approximately 50% identity. Neither new tubulin has greater similarity to the other, more divergent members of the tubulin superfamily, such as γ-, δ-, or ε-tubulin (*31*). Thus, both *Drosophila* and *C. elegans* appear to lack δ- and ε-tubulin, even though δ-tubulin is highly conserved between *Chlamydomonas* and humans. There are also three new members of the central motor domain family of kinesins that encode nonmotor proteins that regulate microtubule dynamics (*32*). There are clear homologs of the dystrophin complex and of dystrobrevin. Finally, the fly lacks cytoplasmic intermediate filament proteins, other than nuclear lamins, although other invertebrates, including *C. elegans*, appear to have genes encoding these (*33*). *Drosophila* and *C. elegans* both also appear to lack a gene encoding kinectin, the proposed receptor for kinesin and cytoplasmic dynein on vesicles and organelles (*34*). Flies and worms must thus use different proteins to link microtubule motors to vesicles and organelles.

Cell adhesion. Cell-cell adhesion and cell-substrate adhesion molecules have been crucial to the development of multicellular organisms and the evolution of complex forms of embryogenesis (*13*). The transmembrane extracellular matrix-cytoskeleton linkage via integrins is ancient. There are five α and two β integrins in the fly, two α and one β in *C. elegans*, and at least 18 α and eight β in vertebrates. Integrin-associated cytoplasmic proteins (talin, vinculin, α-actinin, paxillin, FAK, p130CAS, and ILK) are encoded by single-copy fly genes, as are tensin and syndecan.

Two genes for type IV collagen subunits and genes for the three subunits of laminin were already known in the fly. Analysis of the genome revealed no more laminin genes and only one more collagen, which is closest to types XV and XVIII of vertebrates. A counterpart of this collagen is found in *C. elegans*, which has on the order of 170 collagens. Most important, it appears that the core components of basement membranes (two type IV collagen subunits, three laminin subunits, entactin/nidogen, and one perlecan), are all present in flies. This constitution of basement membranes was clearly established early in evolution and has been well conserved in metazoans; remarkably, the fly preserves the linked head-to-head organization of vertebrate type-IV collagen genes. In contrast to this conservation, many well-known vertebrate integrin (ECM) ligands are absent from the fly: fibronectin, vitronectin, elastin, von Willebrand factor, osteopontin, and fibrillar collagens are all missing.

The fly has three classic cadherins, two of which are closely linked, but no protocadherins of the type found in vertebrates as clusters with common cytoplasmic domains (*35*). Vertebrates have three such clusters encoding over 50 protocadherins and close to 20 classical

cadherins. The fly has no reelin, an ECM ligand for CNR-type protocadherins in vertebrates (*36*). However, there are other fly proteins with cadherin repeats, including the previously known Fat, Dachsous, and Starry night, and a new very large protein related to Fat. *C. elegans* has 15 genes containing cadherin repeats; the number in humans is now 70 and will undoubtedly rise (*13*).

Cell signaling. Components of known signaling pathways in the fly and worm have largely been uncovered by examinations of developmental systems. It is a tribute to the previous genetic analyses done in these organisms that only a modest number of new components of the known signaling pathways were revealed by analysis of the genomic sequence. The core components defined in flies and worms have been used in modified and expanded forms in vertebrates (*37*). The predominant pathways—transforming growth factor–β (TGF-β), receptor tyrosine kinases, Wingless/Wnt, Notch/lin-12, Toll/IL1, JAK/STAT/cytokine, and Hedgehog (HH) signaling networks—all have largely conserved fly and vertebrate components. The worm, by contrast, does not appear to possess the HH or Toll/IL1 pathways, nor does it have all of the components of the Notch/lin-12 network (*38*). Two new proteins of the TGF-β superfamily were identified, bringing the total to seven; all seven are members of the bone morphogenetic protein (BMP) or β-activin subfamilies. We detected no representatives of the other branches of this superfamily, namely the TGF-β, α-inhibin, and Mullerian inhibiting substance (MIS) subfamilies. Three new members of the Wingless/Wnt family were identified, bringing the total to seven. Each of these proteins has sequence similarity to a different vertebrate Wnt protein; this ancient family clearly underwent much of its expansion before the divergence of the arthropod and chordate lineages. There is only one member of the Notch and HH families, in contrast to the many members of these families in vertebrates.

Apoptosis. The core apoptotic machinery of *Drosophila* shares many features in common with that of mammals. Many apoptosis-inducing signals lead to activation of members of the caspase family of proteases. These proteases function in apoptotic processes as cell death signal transducers and death effectors, and in nonapoptotic processes in flies and mammals (*39*). *Drosophila* contains genes encoding 8 caspases, as compared to 4 in the worm and at least 14 in mammals. Three of the fly caspases contain long NH_2-terminal prodomains of 100 to 200 amino acids that are characteristic of caspases that function as signal transducers. These prodomains are thought to mediate caspase recruitment into signaling complexes in which activation occurs in response to oligomerization. In one pathway described in mammals but not in worms, death signals cause the release of proteins, including cytochrome c and the apoptosis-inducing factor (AIF), from mitochondria (*40*). The human protein Apaf-1, in conjunction with cytochrome c, activates CARD domain–containing caspases (*41*). *Drosophila* has an Apaf-1 counterpart, a CARD domain–containing caspase, and AIF; *Drosophila* also has counterparts to the caspase-activated DNAse CAD/CPAN/DFF40, its inhibitor ICAD/DFF45, and the chromatin condensation factor Acinus (*42*).

Pro- and anti-apoptotic BCL2 family members regulate apoptosis at multiple points (*43*). *Drosophila* encodes two BCL2 family proteins, though more divergent family members may exist. Fifteen BCL2 family proteins have been identified in mammals and two in the worm. In addition, inhibitor of apoptosis (IAP) family proteins negatively regulate apoptosis (*44*). They are defined by the presence of one or more NH_2-terminal repeats of a BIR domain, a motif that is essential for death inhibition. *Drosophila* has four proteins with this motif, as compared to seven identified thus far in mammals. There are several BIR domain–containing proteins in *C. elegans* and yeast, but none has been implicated in cell death regulation. Reaper (RPR), Wrinkled (W), and Grim are essential *Drosophila* cell death activators (*45*). Orthologs have not been identified in other organisms, but they are likely to exist because RPR, W, and Grim induce apoptosis in vertebrate systems and physically interact with apoptosis regulators that include IAPs and the *Xenopus* protein Scythe (*46*), for which there is a predicted *Drosophila* homolog.

Neuronal signaling. The neuronal signaling systems in flies, worms, and vertebrates reveal extensive conservation of some components, as well as extreme divergence, or the total absence, of others. There is no voltage-activated sodium channel in the worm (*17*); flies and vertebrates generate sodium-dependent action potentials. The fly genome encodes two pore-forming subunits for sodium channels (Para and NaCP60E), and also four voltage-dependent calcium channel α subunits, including one T-type/α1G, one L-type/α1D (Dmca1D), one N-type/α1A (Dmcα1A), and one protein that is more similar to an outlying *C. elegans* protein than to known vertebrate calcium channels. Additional fly calcium channel subunits include one β, one γ 2, and three α 2 subunits.

The worm genome encodes over 80 potassium channel proteins (*17*); the fly genome has only 30. The extent to which these different family sizes contribute to the establishment of unique electrical signatures is unknown. The fly potassium channel family includes five *Shaker*-like genes (*Shaker*, *Shab*, *Shal*, and two *Shaws*); a large conductance calcium-activated channel gene (*slowpoke*); a slack subunit relative; three members of the *eag* family (*eag*, *sei*, and *elk*); one small conductance calcium-regulated channel gene; one KCNQ channel gene; and four cyclic nucleotide–gated channel genes. In addition, there are 50 TWIK members in the worm, but only 11 fly members of the two-pore/TWIK family with four transmembrane domains. There are also three fly members of the inward rectifier/two transmembrane family. Finally, neither the fly nor the worm has discernible relatives of a number of mammalian channel-associated subunits such as minK and miRP1.

Table 4. The 10 InterPro protein domains occurring in the largest number of different proteins in *S. cerevisiae* and *C. elegans*.

Acc. no.	InterPro domain name	No. of proteins
	S. cerevisiae	
IPR000719	Eukaryotic protein kinase	119
IPR001680	G-protein beta WD-40 repeats	90
IPR001650	DNA/RNA helicase domain (DEAD/DEAH box)	75
IPR001138	Fungal transcriptional regulatory protein, N-terminus	60
IPR001042	TYA transposon protein	57
IPR000504	RNA-binding region RNP-1 (RNA recognition motif)	55
IPR001410	DEAD/DEAH box helicase	48
IPR000822	Zinc finger, C2H2 type	47
IPR001066	Sugar transporter	46
IPR001969	Eukaryotic and viral aspartyl proteases active site	42
	C. elegans	
IPR000168	7-Helix G-protein coupled receptor, nematode (probably olfactory) family	545
IPR000694	Proline-rich region	398
IPR000719	Eukaryotic protein kinase	388
IPR002356	G-protein–coupled receptors, rhodopsin family	335
IPR001628	C4-type steroid receptor zinc finger	224
IPR001810	F-box domain	215
IPR000087	Collagen triple helix repeat	166
IPR001304	C-type lectin domain	165
IPR002900	Domain of unknown function	142
IPR000822	Zinc finger, C2H2 type	138

There are also major differences postsynaptically. *C. elegans* has approximately 100 members of a family of ligand-gated ion channels (*17*); flies have about 50. The worm has 42 nicotinic acetylcholine receptor subunits and 37 GABA(A)-like receptor subunits; the fly contains only 11 nicotinic receptor subunit genes and 12 GABA(A)/glycine-like receptor subunit genes. In contrast, there are 30 members of the excitatory glutamate receptor family in the fly but only 10 in the worm. These include subtypes of the AMPA, kainate, NMDA, and delta families. In addition, the fly genome contains a large number of PDZ-containing genes, approximately a dozen of which encode proteins that have high sequence similarity to mammalian proteins that interact with specific subsets of ion channels. We also found a number of additional ion channel families, including three voltage-dependent chloride channels, 14 Trp-like channels, 24 amiloride-sensitive/degenerin-like sodium channels, one ryanodine receptor, one IP_3 (inositol 1,4,5-trisphosphate) receptor, eight innexins, and two porins. *C. elegans* is missing a nitric oxide synthase gene, copies of which occur in fly and vertebrate genomes.

A large array of proteins mediates specific aspects of synaptic vesicle trafficking and contributes to the conversion of electrical signals to neurotransmitter release. These components of exocytosis and endocytosis are relatively well conserved with respect to both domain structures and amino acid identities (50 to 90%). The fly has enzymes for the synthesis of the neurotransmitters glutamate, dopamine, serotonin, histamine, GABA, acetylcholine, and octopamine, and a family of conserved transporters is likely to be involved in loading vesicles with these neurotransmitters. The conserved vesicular trafficking proteins, with 50 to 80% amino acid identity, include members of the Munc-18, SCAMP, synaptogyrin, HRS2, tomosyn, cysteine string protein, exocyst (SEC 5, 6, 7, 8, 10, 13, 15, EXO 70, and EXO84), synapsin, rabphilin-3A, RIM, rab-3, CAPS, Mint, Munc-13, NSF, α and γ SNAP, DOC-2B, latrophilin, Veli, CASK, VAP-33, Snapin, SV2, and complexin families. Generally, there is only one homolog in *Drosophila* for every three to four isoforms in mammals. However, there are eight fly synaptotagmin-like genes, making this the largest family of vesicle proteins in *Drosophila* (*47*). However, there is no homolog of synaptophysin, an early candidate for a vesicle fusion pore, which indicates a nonessential role in exocytosis for this particular protein across phyla.

Membrane trafficking also requires interactions between compartment-specific vesicular and target membrane proteins (v-SNAREs and t-SNAREs, respectively), whose subcellular distribution and combinatorial binding patterns are predicted to define organelle identity and targeting specificity (*48*). The completed fly genome allows us to address whether there is any correlation between the increased developmental complexity of multicellular organisms and a larger number of SNAREs than that found in unicellular organisms. In the fly, we find six synaptobrevins, three SNAP-25s, 10 syntaxins, and four additional t-SNAREs (membrin, BET1, UFE1, and GOS28), and the number of SNAREs is similar between yeast (*49*) and *Drosophila*. Thus, basic subcellular compartmentalization and membrane trafficking to and between these various compartments has not changed dramatically in multicellular versus unicellular organisms. Dynamin, clathrin, the clathrin adapter proteins, amphiphysin, synaptojanin, and a number of additional genes that encode proteins with defined endocytotic motifs are all present.

In contrast to the conservation of the synaptic vesicle trafficking machinery, the few identified proteins present at mammalian active zones, namely aczonin, bassoon, and piccolo, do not have relatives in *Drosophila*. There are, however, numerous proteins in the fly with combinations of C2 domains, PDZ domains, zinc fingers, and proline-rich domains, indicating that the precise protein composition of active zones is likely to vary among metazoans. In addition, *Drosophila* contains a neurexin III gene and four neuroligin genes that may be part of a neurexin-neuroligin complex that has been widely proposed to provide a synaptic scaffold for linking pre- and postsynaptic structures in mammals (*50*). Potential agrin and Musk genes are also present, though the overall sequence similarity is low.

Immunity. Multicellular organisms have elaborate systems to defend against microbial pathogens. Only vertebrates have an acquired immune system, but both vertebrates and invertebrates share a more primitive innate immune system. Innate immunity is based on the detection of common microbial molecules such as lipopolysaccharides and peptidoglycans by a class of receptors known as pattern recognition receptors (*51*). We identified a large family of genes encoding homologs of receptors that are involved in microbial recognition in other organisms. These include two new homologs of the *Drosophila* Scavenger Receptors (dSR-CI), nine members of the CD36 family, 11 members of the peptidoglycan recognition protein (PGRP) family, three Gram-negative binding protein (GNBP) homologs, and several lectins (*52*).

The recognition of infection by immunoresponsive tissues induces a battery of defense genes via Toll/nuclear factor kappa B (NF-κB) pathways in both *Drosophila* and mammals (*53*). The Toll receptor was initially discovered as an essential component of

Table 5. Proteins in *D. melanogaster*, *C. elegans*, and *S. cerevisiae* with more than one InterPro domain. These numbers represent the total number of recognizable domains within a single protein, no matter whether they are multiple copies of the same domain or different domains.

InterPro domains per protein	*D. melanogaster* (number of proteins)	*C. elegans* (number of proteins)	*S. cerevisiae* (number of proteins)
2	920	1236	410
3	388	458	121
4	219	182	58
5	163	98	26
6	101	72	17
7	92	53	15
8	58	27	7
9	42	25	4
10	22	18	7
11–15	73	43	6
16–20	18	17	1
21–30	22	22	0
31–50	8	5	0
51–75	4	5	0

Table 6. Proteins in *D. melanogaster*, *C. elegans*, and *S. cerevisiae* with multiple different InterPro domains. Individual InterPro domains are counted only once per protein, regardless of how many times they occur in that protein.

Unique InterPro domains per protein	*D. melanogaster* (number of proteins)	*C. elegans* (number of proteins)	*S. cerevisiae* (number of proteins)
2	1474	1248	402
3	413	335	95
4	156	114	23
5	52	38	4
6	8	9	1
7 or more	4	3	0

the pathway that establishes the dorsoventral axis of the *Drosophila* embryo. Recent genetic studies now reveal that Toll signaling pathways are key mediators of immune responses to fungi and bacteria in both *Drosophila* and mice (*53*). We found seven additional homologs of Toll proteins in *Drosophila*, all of which are more similar to each other than to their mammalian counterparts. Some of these other Toll proteins, like 18-wheeler, will probably mediate innate immune responses. In *Drosophila*, infection by at least some microbes induces a proteolytic cascade that leads to the processing of Spaetzle (SPZ), a cytokine-like protein, which then activates Toll (*53*). We found two proteins related to SPZ with similarities that include most or all of the cysteine residues of SPZ. Given the presence of multiple Toll-like receptors in *Drosophila*, these new SPZ-like proteins may also function in the immune system. With the exception of the two I-κB kinase homologs and the three rel proteins (Dorsal, Dif, and Relish), the *Drosophila* genome appears to contain only single copies of the genes encoding intracellular components of the Toll pathway: Tube, Pelle, and Cactus. How do the different Toll receptors trigger specific immune responses using the same intracellular intermediates? One explanation is that additional signaling components remain unidentified; another explanation is crosstalk with other signaling pathways. In contrast, a Toll ortholog has not been identified in *C. elegans*, although there are some Toll-like receptors. *C. elegans*, in addition, does not possess homologs of NF-κB/dorsal transcriptional activators that function downstream of Toll. Although it is probable that the worm has retained parts of the innate immunity network, there is no clear evidence of an inducible host defense system in the worm.

One of the most potent innate immune responses in insects is the transcriptional induction of genes encoding antimicrobial peptides (*53*). In contrast to Metchnikowin, Drosocin, and Defensin peptides, which are encoded by single genes, the sequence data indicate that, like the previously identifed cecropin clusters, several antimicrobial peptides are encoded by gene families that are larger than previously suspected. Four genes appear to encode antifungal peptide Drosomycin isoforms, and two genes each code for the antibacterial proteins Attacin and Diptericin. These additional genes may generate peptides with slightly different spectra of antimicrobial activity or may simply amplify the antimicrobial response.

Concluding Remarks

What have we learned about the proteins encoded by the three sequenced eukaryotic genomes? Some information emerges readily from the comparison of the fly, worm, and yeast genomes. First, the core proteome sizes of flies and worms are similar and are only twice the size of that of yeast. This is perhaps counterintuitive, because the fly, a multicellular animal with specialized cell types, complex development, and a sophisticated nervous system, looks more than twice as complicated as single-celled yeast. The lesson is that the complexity apparent in the metazoans is not achieved by sheer number of genes (*54*). Second, there has been a proliferation of bigger and more complex proteins in the two metazoans relative to yeast, including, not surprisingly, more proteins with extracellular domains involved in cell-cell and cell-substrate interactions. Finally, the population of multidomain proteins is somewhat larger and more diverse in the fly than in the worm. There is presently no practical way to quantify differences in biological complexity between two organisms, however, so it is not possible to correlate this increased domain expansion and diversity in the fly with differences in development and morphology.

The availability of the annotated sequence of the *Drosophila* genome enhances the fly's usefulness as an experimental organism. By greatly facilitating positional cloning, the genome sequence will increase the efficiency of genetic screens that seek to identify genes underlying many complex processes of cell biology, development, and behavior. Such screens have been the mainstay of *Drosophila* research and have contributed enormously to our knowledge of metazoan biology. The genome sequencing effort has revealed a number of previously unknown counterparts to human genes involved in cancer and neurological disorders; for example, *p53*, *menin*, *tau*, *limb girdle muscular dystrophy type 2B*, *Friedrich ataxia*, and *parkin*. All of these fly genes are present in a single copy in the genome and can be genetically analyzed without uncertainty about redundant copies. More genetic screens are important in order to uncover interacting network members. Orthologs of these network members can then be sought in the human genome to determine if alterations in any of them predispose humans to the disease in question, an experimental paradigm that has already been successfully executed in several cases. Flies can also play an important role in exploring ways to rectify disease phenotypes. For example, at least 10 human neurodegenerative diseases are caused by expansion of polyglutamine repeats (*55*). Human proteins containing expanded polyglutamine repeats have been expressed in flies, resulting in the formation of nuclear inclusions that contain the protein as well as other shared components (*56*), just as in humans. It has been shown that directed expression of the human HSP70 chaperone in the fly can totally suppress neurodegeneration resulting from expression of the human spinocerebellar ataxia type 3 protein (*57*). The power and speed of this in vivo system are unparalleled, and we anticipate the increased use of such "humanized" fly models.

Knowing the complete genomic sequence also allows new experimental approaches to long-standing problems. For example, it makes it possible to study networks of genes rather than individual genes or pathways. Assaying the level of transcription of every gene in the genome makes it at least theoretically possible to monitor the expression of an entire network of genes simultaneously. One problem that is approachable this way is the combinatorial control of gene transcription. The fly genome appears to encode only about 700 transcription factors, and mutations in over 170 have already been isolated and characterized. The techniques are available to measure the changes in expression of every gene in individual cell types as a consequence of loss or overexpression of each transcription factor. We can look for common sequence elements in the promoters of coregulated genes and perform chromatin immunoprecipitation to identify the in vivo binding sites of individual factors. For the first time, we can envision obtaining the data needed to understand the behavior of a complex regulatory network. Of course, collecting these data is a massive task, and developing methods to analyze the data is even more daunting. But it is no longer ludicrous to try.

How big is the core proteome of humans? Vertebrates have many gene families with three or four members: the HOX clusters, calmodulins, Ezrins, Notch receptors, nitric oxide synthases, syndecans, and NF1 transcription factor genes are some examples (*58*). This is evidence for two genome doublings during mammalian evolution, superimposed on which were the amplifications and contractions over evolutionary time that uniquely characterize each lineage (*59*). The human genome, with 80,000 or so genes, is likely to be an amplified version of a very much smaller genome, and its core proteome may not be much larger than that of the fly or worm; that is, the more complex attributes of a human being are achieved using largely the same molecular components. The evolution of additional complex attributes is essentially an organizational one; a matter of novel interactions that derive from the temporal and spatial segregation of fairly similar components.

Finally, approximately 30% of the predicted proteins in every organism bear no similarity to proteins in its own proteome or in the proteomes of other organisms. In other words, sequence similarity comparisons consistently fail to give us information about nearly a third of the components that make every organism uniquely itself. What does this mean with respect to the evolution and function of these proteins? Does each genome contain a subpopulation of very rapidly evolving genes? One-third of randomly chosen cDNA clones do

not cross-hybridize between *D. melanogaster* and *Drosophila virilis* (*60*). Even though these are distantly related species, they are developmentally and morphologically very similar. Crystallographic data will be needed to determine whether these proteins that have diverged in primary sequence have maintained their three-dimensional structures or have diverged so far that new folds and domains have formed.

Our first look at the annotated fly genome provokes these and other questions. Access to the genomic sequence will help us design the experiments needed to answer them. The relative simplicity and manipulability of the fly genome means that we can address some of these biological questions much more readily than in vertebrates. That is, after all, what model organisms are for.

References and Notes

1. M. D. Adams *et al.*, *Science* **287**, 2185 (2000); *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998); A. Goffeau *et al.*, *Science* **274**, 546 (1996).
2. R. D. Fleischman *et al.*, *Science* **269**, 496 (1995).
3. *C. elegans* data were taken from A C. Elegans Database (ACEDB) release WS8.
4. Local gene duplications were determined by searching for *N* similar genes within 2*N* genes on each arm. For example, if three similar genes are found within a region containing six genes, this counts as one cluster of three genes. Genes were judged to be similar if a BLASTP High Scoring Pair (HSP) with a score of 200 or more existed between them. Histone gene clusters were not included. *C. elegans* data were taken from ACEDB release WS8, containing 18,424 genes.
5. More information about GO is available at http://www.geneontology.org/. The Gene Ontology project provides terms for categorizing gene products on the basis of their molecular function, biological role, and cellular location using controlled vocabularies.
6. Initial results came from an NxN BLASTP analysis performed for each fly, worm, and yeast sequence in a combined data set of these completed proteomes. The databases used are as follows: Celera–Berkeley *Drosophila* Genome Project (BDGP), 14,195 predicted protein sequences (1/5/2000); WormPep 18, Sanger Centre, 18,576 protein sequences; and Saccharomyces Genome Database (SGD), 6306 protein sequences (1/7/2000). A version of NCBI-BLAST2 was used with the SEG filter and with the effective search space length (Y option) set to 17,973,263. Pairs were formed between every query sequence with a significant BLASTP to one of the other organisms' sequences. Significance was based on E-value cutoffs and length of match. These pairs were then independently grouped using single linkage clustering (*61*). Finally, the number of proteins from each proteome was counted. The requirement for 80% alignment of sequences makes this method of defining orthology particularly sensitive to errors that arise from incorrect protein prediction. However, the results comparing yeast and worm are essentially identical to those previously reported (*61*), even though the effective database size was different, the data sets have changed (Chervitz: yeast 6217 and worm 19,099; this study: yeast 6306, and worm 18,576), and the version of BLAST used is quite different (Chervitz: WashU BLAST 2.0a19MP; this study: NCBI BLAST 2.08).
7. A. Bairoch and R. Apweiler, *Nucleic Acids Res.* **28**, 45 (2000).
8. J. G. Henikoff, E. A. Greene, S. Pietrokovski, S. Henikoff, *Nucleic Acids Res.* **28**, 228 (2000).
9. InterPro (Integrated resource for protein domains and functional sites) is a collaborative effort of the SWISS-PROT, TrEMBL, PROSITE, PRINTS, Pfam, and ProDom databases to integrate the different pattern databases into a single resource. The database and a detailed description of the project can be found under http://www.ebi.ac.uk/interpro/. PROSITE is described in K. Hofmann, P. Bucher, L. Falquet, A. Bairoch, *Nucleic Acids Res.* **27**, 215 (1999); PFAM is described in A. Bateman *et al.*, *Nucleic Acids Res.***27**, 260 (1999); and PRINTS is described in T. K. Attwood *et al.*, *Nucleic Acids Res.* **27**, 220 (1999).
10. G. D. Plowman, S. Sudarsanam, J. Bingham, D. Whyte, T. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13603, (1999).
11. J. Barrett, N. D. Rawlings, J. F. Wessner, Eds., *Handbook of Proteolytic Enzymes* (Academic Press, San Diego, CA, 1998).
12. C. L. Smith and R. DeLotto, *Nature* **368**, 548 (1994); K. D. Konrad, T. J. Goralski, A. P. Mahowald, J. L. Marsh, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6819 (1998); E. K. LeMosy, C. C. Hong, C. Hashimoto, *Trends Cell Biol.* **9**, 102 (1999).
13. R. O. Hynes, *Trends Cell Biol.* **9**, M33 (1999).
14. P. Bork, A. K. Downing, B. Kieffer, I. D. Campbell, *Quart. Rev. Biophys.* **29**, 119 (1996).
15. P. Vernier, B. Cardinaud, O. Valdenaire, H. Philippe, J.-D. Vincent, *Trends Pharmacol. Sci.* **16**, 375, (1995); J. Colas, J. Launay, J. Vonesch, P. Hickel, L. Maroteaux, *Mech. Dev.* **87**, 77 (1999); M. R. Costa, E. T. Wilson, E. Wieschaus, *Cell* **76**, 1075 (1994).
16. P. Mombaerts, *Science* **286**, 707, (1999).
17. C. I. Bargmann, *Science* **282**, 2028 (1998).
18. P. J. Clyne *et al.*, *Neuron* **22**, 327 (1999); L. B. Vosshall, H. Amrein, P. S. Morozov, A. Rzhetsky, R. Axel, *Cell* **96**, 725 (1999); P. P. Laissue *et al.*, *J. Comp. Neurol.* **405**, 543 (1999).
19. Y. J. Lin, L. Seroude, S. Benzer, *Science* **282**, 943 (1998).
20. Y. Zhang, Y. Xiong, W. G. Yarbrough, *Cell* **92**, 725 (1998).
21. S. N. Jones, A. E. Roe, L. A. Donehower, A. Bradley, *Nature* **378**, 206 (1995).
22. I. The *et al.*, *Science* **276**, 791 (1997).
23. N. Ito and G. M. Rubin, *Cell* **96**, 529 (1999).
24. M. O. Hengartner and H. R. Horvitz, *Cell* **76**, 665 (1994).
25. F. Hauser, H. P. Nothacker, C. J. Grimmelikhuijzen, *J. Biol. Chem.* **272**, 1002 (1997).
26. P. R. Mueller, T. R. Coleman, A. Kumagai, W. G. Dunphy, *Science* **270**, 86 (1995).
27. B. D. Dynlacht, A. Brook, M. Dembski, L. Yenush, N. Dyson, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6359 (1994); W. Du, M. Vidal, J.-E. Xie, N. Dyson, *Genes Dev.* **10**, 1206 (1996); T. Sawado *et al.*, *Biochem. Biophys. Res. Commun.* **251**, 409 (1998).
28. X. Lu and H. R. Horvitz, *Cell* **95**, 981 (1998).
29. T. Kreis and R. Vale, Eds., *Guidebook to the Cytoskeletal and Motor Proteins* (Oxford Univ. Press, Oxford, 1999).
30. P. Chang and T. Stearns, *Nature Cell Biol.* **2**, 30 (2000).
31. S. K. Dutcher and E. C. Trabuco, *Mol. Biol. Cell* **9**, 1293 (1998).
32. A. Desai, S. Verma, T. J. Mitchison, C. E. Walczak, *Cell* **96**, 69 (1999).
33. K. Weber, in (*29*), pp. 291–293.
34. J. Kumar, H. Yu, M. P. Sheetz, *Science* **267**, 1834 (1995).
35. Q. Wu and T. Maniatis, *Cell* **97**, 779 (1999).
36. K. Senzaki, M. Ogawa, T. Yagi, *Cell* **99**, 635 (1999).
37. M. P. Belvin and K. V. Anderson, *Annu. Rev. Cell Dev. Biol.* **12**, 393 (1996); M. Hammerschmidt, A. Brook, A. P. McMahon, *Trends Genet.* **13**, 14 (1997); C. M. Blaumueller and S. Artavanis-Tsakonas, *Perspect. Dev. Neurobiol.* **4**, 325 (1997); T. Hunter, *Philos. Trans. R. Soc. London Ser. B* **353**, 583 (1998); K. M. Cadigan and R. Nusse, *Genes Dev.* **11**, 3286 (1997); J. Capdevila and J. C. Belmonte, *Curr. Opin. Genet. Dev.* **9**, 427 (1999); L. Engstrom, E. Noll, N. Perrimon, *Curr. Top. Dev. Biol.* **35**, 229 (1997); B. E. Stronach and N. Perrimon, *Oncogene* **18**, 6172 (1999); P. W. H. Holland, J. Garcia-Fernandez, N. A. Williams, A. Sidow, *Development* (suppl.) (1994), p. 125.
38. G. Ruvkun and O. Hobert, *Science* **282**, 2033 (1998).
39. W. C. Earnshaw, L. M. Martins, S. H. Kaufmann, *Annu. Rev. Biochem.* **68**, 383 (1999); A. Zeuner, A. Eramo, C. Peschle, R. DeMaria, *Cell Death Diff.* **6**, 1075 (1999).
40. X. Liu, C. N. Kim, J. Yang, R. Jemmerson, X. Wang, *Cell* **86**, 147 (1996); S. A. Susin *et al.*, *Nature* **397**, 441 (1999).
41. P. Li *et al.*, *Cell* **91**, 479 (1997).
42. A. G. Park, *Trends Cell Biol.* **10**, 394 (2000); S. Sahara *et al.*, *Nature* **401**, 168 (1999).
43. A. Gross, J. M. McDonnell, S. J. Korsmeyer, *Genes Dev.***13**, 1899 (1999).
44. L. K. Miller, *Trends Cell Biol.* **9**, 323 (1999).
45. J. M. Abrams, *Trends Cell Biol.* **9**, 435 (1999).
46. K. Thress, W. Henzel, W. Shillinglaw, S. Kornbluth, *EMBO J.* **17**, 6135 (1998).
47. J. T. Littleton, T. L. Serano, G. M. Rubin, B. Ganetzky, E. R. Chapman, *Nature* **400**, 757 (1999).
48. T. Solner *et al.*, *Nature* **362**, 318 (1993).
49. R. Jahn and T. C. Sudhof, *Annu. Rev. Biochem.* **68**, 863 (1999).
50. K. Ichtchenko *et al.*, *Cell* **81**, 435 (1995).
51. R. Medzhitov and C. A. Janeway Jr., *Cell* **91**, 295 (1997).
52. A. Pearson, *Current Opin. Immunol.* **8**, 20 (1996); N. C. Franc *et al.*, *Immunity* **4**, 431 (1996); D. Kang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10078 (1998); W. J. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7888 (1996).
53. J. A. Hoffmann and J.-M. Reichhart, *Trends Cell Biol.* **7**, 309 (1997); K. V. Anderson, *Curr. Opin. Immun.***12**, 13 (2000).
54. G. L. G. Miklos, *J. Am. Acad. Arts Sci.* **127**, 197 (1998).
55. M. Perutz, *Trends Biochem. Sci* **24**, 58 (1999).
56. J. M. Warrick *et al.*, *Cell* **93**, 939 (1998); G. R. Jackson *et al.*, *Neuron* **21**, 633 (1998).
57. J. M. Warrick *et al.*, *Nature Genet.* **23**, 425 (1999).
58. J. Spring, *FEBS Lett.***400**, 2 (1997).
59. S. Aparicio, *Trends Genet.* **16**, 54 (2000).
60. K. J. Schmid and D. Tautz, *Proc. Natl. Acad. Sci. USA.* **94**, 9746 (1997).
61. S. A. Chervitz *et al.*, *Science* **282**, 2022 (1998).
62. See www.sciencemag.org/feature/data/1049664.shl for complete protein domain analysis.
63. Paralogous gene families (Table 1) were identified by running BLASTP. A version of NCBI-BLAST2 optimized for the Compaq Alpha architecture was used with the SEG filter and the effective search space length (Y option) set to 17,973,263. Each protein was used as a query against a database of all other proteins of that organism. A clustering algorithm was then used to extract protein families from these BLASTP results. Each protein sequence constitutes a vertex; each HSP between protein sequences is an arc, weighted by the BLAST Expect value. The algorithm identifies protein families by first breaking all arcs with an E value greater than some user-defined value (1×10^{-6} was used for all of the analyses reported here). The resulting graph is then split into subgraphs that contain at least two-thirds of all possible arcs between vertices. The algorithm is "greedy"; that is, it arbitrarily chooses a starting sequence and adds new sequences to the subgraph as long as this criterion is met. An interesting property of this algorithm is that it inherently respects the multidomain nature of proteins: For example, two multidomain proteins may have significant similarity to one another but share only one or a few domains. In such a case, the two proteins will not be clustered if the unshared domains introduce a large number of other arcs.
64. An NxN BLASTP analysis was performed for each fly, worm, and yeast sequence in a combined data set of these completed proteomes. The databases used are as follows: Celera-BDGP, 14,195 predicted protein sequences (1/5/2000); WormPep18, Sanger Centre, 18,424 protein sequences; and SGD, 6246 protein sequences (1/7/2000). BLASTP analysis was also performed against known mammalian proteins (2/1/2000, GenBank nonredundant amino acid, Human, Mouse, and Rat, 75,236 protein sequences), and TBLASTN analysis was performed against a database of mammalian ESTs (2/1/00, GenBank dbEST, Human, Mouse, and Rat). A version of NCBI-BLAST2 optimized for the Compaq Alpha architecture was used with the SEG filter and the effective search space length (Y option) set to 17,973,263.

65. The many participants from academic institutions are grateful for their various sources of support. Participants from the Berkeley *Drosophila* Genome Project are supported by NIH grant P50HG00750 (G.M.R.) and grant P4IHG00739 (W.M.G.).

EXPERIMENTAL CELL RESEARCH 213, 183–190 (1994)

Purification and Characterization of Integrin $\alpha 9\beta 1$

E. FORSBERG,* B. EK,† Å. ENGSTRÖM,‡ AND S. JOHANSSON*,[1]

*Department of Medical and Physiological Chemistry, University of Uppsala, The Biomedical Center, Box 575, S-751 23 Uppsala, Sweden; †Department of Cell Research, Swedish University of Agricultural Sciences, Uppsala Genetic Center, Box 7055, S-750 07 Uppsala, Sweden; and ‡Department of Immunology, University of Uppsala, The Biomedical Center, Box 582, S-751 23 Uppsala, Sweden

A new $\beta 1$-containing integrin was isolated from rat liver by affinity chromatography on Sepharose conjugated with the peptide GRGDSPC. The interaction was weakened but not abolished when the arginine and/or aspartic acid in the peptide were replaced with lysine and glutamic acid, respectively. In contrast, the cysteine was necessary for binding of the integrin. The $\beta 1$-associated protein, referred to as $\alpha 9$, had an N-terminal amino acid sequence related to but distinct from previously described integrin α-subunits. In addition, an internal peptide sequence was obtained which confirmed that the protein is a new member of the family of integrin α-subunits. An antiserum raised against a synthetic peptide corresponding to amino acids 1–16 of $\alpha 9$ reacted specifically with this protein and was used to identify $\alpha 9$ in several tissues. The integrin $\alpha 9\beta 1$ was not retained on Sepharose conjugated with Englebreth–Holm–Swarm tumor (EHS)–laminin, collagen type I, or a 105-kDa cell-binding fragment of fibronectin. However, it did bind specifically to EHS–laminin and collagen type I adsorbed to plastic microtiter wells. The sites of the interactions were localized to fragment E8 of EHS–laminin and to cyanogen bromide fragment 8 of collagen $\alpha 1$(I) and were not inhibited by soluble RGD-containing peptides. The results indicate that $\alpha 9\beta 1$ is a widely distributed laminin/collagen receptor which may have additional, yet unidentified ligands. © 1994 Academic Press, Inc.

INTRODUCTION

The integrins are a family of heterodimeric cell surface glycoproteins which mediate interactions between cells and the extracellular matrix (ECM) and in some cases take part in cell–cell adhesion [1–3]. These receptors are composed of noncovalently associated α- and β-subunits, which both span the plasma membrane. So far, 13 α- and 8 β-subunits have been described, which are capable of forming at least 19 different heterodimers. In addition, several variants generated by alternative splicing of the mRNAs are known. This multiplicity of ECM receptors allows cells to recognize and respond to alterations in the composition of ECMs. The cytoplasmic domains of the integrin subunits have been strongly conserved during evolution, indicating that they interact with other conserved structures inside the cells. Since all of the known integrin subunits have distinct cytoplasmic domains, they may have different functions, e.g., be parts of separate signaling pathways, or be targets for regulation by different mechanisms.

A number of integrins are known to interact with the tripeptide Arg–Gly–Asp (RGD). This motif was discovered as a cell-binding structure in fibronectin (FN) [4] and has later been found in many other proteins of the ECM. Substitution of single amino acids in the tripeptide usually abolishes binding of receptors [5], although variants of the peptide have been shown to be functional in some cases [6, 7]. In addition, the amino acids surrounding the RGD sequence are of importance, influencing the specificity and affinity for the receptors [8–12]. The detailed structural requirements for the interactions between integrins and RGD-containing proteins are not fully understood. One interesting suggesting is that the aspartic acid in the RGD motif may be involved in the coordination of divalent cations which are known to be necessary for binding of integrins to their ligands [13].

Several cell adhesion proteins of the extracellular matrix, e.g., FN, laminin (LN), and collagen, seem to possess RGD-dependent as well as RGD-independent binding sites for integrins [14–19]. Conversely, some integrins recognize both RGD-containing determinants and protein structures lacking an RGD sequence [20, 21]. In this study, a novel integrin was isolated which seems to have such a dual ligand specificity.

MATERIALS AND METHODS

Proteins and peptides. FN was purified from human plasma as previously described [22]. FN was digested with chymotrypsin and the cell-binding fragment (105 kDa) was isolated as described [23]. Vitronectin (VN) was a gift from Dr. Björn Dahlbäck (Lund, Sweden), fibrinogen was from KABI-Pharmacia (Stockholm), and collagen type I was from Vitrogen. The collagen $\alpha 1$(I) derived CNBr fragments CB3, CB7, and CB8 were gifts from Dr. Kristofer Rubin (Upp-

[1] To whom correspondence and reprint requests should be addressed. Fax: (46)-18-550762.

0014-4827/94 $6.00

sala, Sweden). LN, isolated from the Engelbreth-Holm-Swarm (EHS) tumor, and its fragments E8 and P1 [24], were gifts from Dr. Mats Paulsson (Bern, Switzerland). The peptides GRGDSPC, GRGESPC, GKGDSPC, GKGESPC, GRGDSP, and YNIDAQRPVRFQGPPGC ($\alpha 9_{1-16}$) were synthesized by Fmoc chemistry utilizing activation with HBTU [25]. The peptides were deprotected by reagent K [26], precipitated with diethylether, and dried under vacuum. The peptides were analyzed by plasma desorbtion mass spectrometry to verify the predicted masses and used without further purification. Peptide GRGDS was obtained from Calbiochem-Novabiochem. $\alpha 9_{1-16}$ corresponds to the N-terminal amino acids 1–16 of the integrin subunit $\alpha 9$ (see Fig. 3A) with an additional cysteine at position 17.

Proteins were coupled to CNBr-activated Sepharose and peptides were coupled to activated CH-Sepharose in 0.1 *M* $NaHCO_3$, pH 8.0, 0.15 *M* NaCl according to the recommendations of the manufacturer (Pharmacia-LKB Technology). During the coupling procedure of cysteine-containing peptides to CH-Sepharose, samples were continuously taken from parallel incubations of peptides in 0.1 *M* NH_4CO_3, pH 8.0, and analyzed by mass spectrometry to monitor dimerization of peptides due to oxidation of the thiol group. No dimerization of the peptides in these experiments was detected even after prolonged incubations (90 min) at room temperature (not shown). Further, a sample of the GRGDSPC-Sepharose was eluted with reducing agent (1 mg/ml DTT). No peptides were detected in the DTT eluate by mass spectrometry.

Antibodies. The antisera against the $\beta 1$- and $\alpha 5$-subunits have been described previously [27, 28]. The rabbit antisera against the integrin subunits $\beta 3$ and αv were kind gifts from Dr. Åke Oldberg (Lund, Sweden) and Dr. James Gailit (La Jolla, CA), respectively. An antiserum against $\alpha 9$ was produced by immunizing a rabbit with the $\alpha 9_{1-16}$ peptide. Each injection, given intramuscularly at 2-week intervals, contained Freund's adjuvant mixed with 150 μg of free $\alpha 9_{1-16}$ and 350 μg of ovalbumin conjugated with the peptide by maleimidobenzoyl-*N*-hydrosuccinimide as described [29].

Electrophoresis and immunoblotting. Polyacrylamide gel electrophoresis in sodium dodecyl sulfate (SDS-PAGE) was performed in 7% gels [30]. Immunoblot analysis of proteins electrophoretically transferred to nitrocellulose after SDS-PAGE was performed as described [31]. The nitrocellulose filters were incubated with antiserum at a 1:50 dilution and the specifically bound antibodies were allowed to react with ^{125}I-labeled protein A. In some experiments, the enhanced chemiluminescence method was used instead. In these cases, the protein A step was replaced by incubation with peroxidase-conjugated anti-rabbit IgG at a 1:5000 dilution under conditions recommended by the manufacturer (Amersham). For detection of recognized antigens by either method, the nitrocellulose filters were exposed to X-ray films (Fuji). The molecular weights of the blotted proteins were estimated by relating their positions to those of marker proteins transferred from the polyacrylamide gels.

Purification of integrins. Seven to ten rat livers were homogenized on ice in 200 ml of a buffer containing 10 m*M* Tris/HCl, pH 7.4, 10 m*M* EDTA, 2% Triton X-100, 2 m*M* PMSF, and 1 mg/ml pepstatin A. The homogenate was centrifuged at 25,000*g* for 45 min and the supernatant was applied at a flow rate of 50 ml/h to a wheat germ agglutinin (WGA)-Sepharose (50 ml bed volume, 250 mg WGA). The column was washed with a buffer containing 10 m*M* Tris/HCl, pH 7.4, 50 m*M* NaCl, 0.2% Triton X-100, 0.2 m*M* PMSF, 1 mg/ml pepstatin A, and 2 m*M* $MnCl_2$ (column buffer) and eluted in the same buffer containing 0.3 *M* *N*-acetylglucosamine. The eluted proteins were applied at 5 ml/h to columns of Sepharose conjugated with the 105-kDa FN fragment (10 mg coupled to 1.5 ml of Sepharose), collagen I (20 mg coupled to 3 ml of Sepharose), EHS-LN (10 mg coupled to 1.5 ml of Sepharose), or different RGD-containing peptides (4 mg of peptides coupled to 0.6 ml of Sepharose). After washing, integrins were eluted off the columns with 10 m*M* EDTA in 10 m*M* Tris/HCl, pH 7.4, 150 m*M* NaCl, 0.2% Triton X-100, 0.2 m*M* PMSF, and 1 mg/ml pepstatin A. All steps after the tissue homogenization were performed at 4°C. The eluted integrins were analyzed by SDS-PAGE. For ^{125}I-labeling of the integrins, the chloramine-T method using Iodobeads (Pierce Chemical Co.) was applied.

Amino acid sequence determination. Amino-terminal sequencing of isolated integrins was performed after separation of the α- and β-subunits in 5% SDS-PAGE and transfer of the proteins to PVDF membranes (Millipore) [32]. To prevent blockage of the amino terminal during electrophoresis, thioglycolate was added to the gel solution (0.1 m*M*), to the upper electrophoresis buffer (0.05 m*M*), and to the sample buffer (4 m*M*). After staining of the PVDF membrane with Coomassie BB, the bands were cut out and subjected to amino acid sequencing in a gas-phase sequenator (Applied Biosystems Model 476). For internal amino acid sequencing, the integrin subunit was incubated with a lysine-specific enzyme from Achromobacter Lyticus followed by separation of obtained fragments by reverse-phase chromatography on a C-4 column (2.1 × 30 mm, Aquapore, Brownlee). The repetitive yield was between 92 and 95% for all reactions of the amino acid sequences presented.

Solid-phase receptor assay. Wells of plastic ELISA microtiter plates (Dynatech Laboratories, Inc.) were coated with the indicated proteins in 150 μl coating buffer (10 m*M* Tris, pH 7.4, 150 m*M* NaCl, 1 m*M* $CaCl_2$, 1 m*M* $MgCl_2$) overnight at 4°C. Subsequently, the wells were incubated with the same buffer containing 1.5% bovine serum albumin (BSA) for 2 h at 22°C and washed twice with binding buffer (10 m*M* Tris, pH 7.4, 50 m*M* NaCl, 1 m*M* $CaCl_2$, 1 m*M* $MgCl_2$, 0.1% Triton X-100). The ^{125}I-labeled integrin was incubated in the wells in binding buffer for 90 min at room temperature. After washing with binding buffer, bound receptors were released by the addition of 1 ml 1% SDS and quantified in a gamma counter.

Cell attachment assay. Hepatocytes were isolated by collagenase perfusion of the liver and used for cell attachment studies as described [17].

RESULTS

Purification and Identification of Integrin Dimers

By use of affinity chromatography, one major FN receptor, integrin $\alpha 5\beta 1$, and one collagen/LN receptor, integrin $\alpha 1\beta 1$, have previously been identified on rat hepatocytes [14, 17, 19]. While the integrin $\alpha 5\beta 1$ recognizes the RGD sequence in FN [5, 14], the integrin $\alpha 1\beta 1$ binds to collagen type I and EHS-LN independently of RGD sites. By cell attachment experiments there were indications for additional LN receptors on hepatocytes, some of which appeared to recognize RGD (Forsberg and Johansson, unpublished observations). In attempts to isolate RGD-binding proteins from rat liver, the solubilized tissue was applied to Sepharose conjugated with different RGD-containing peptides. When a synthetic peptide consisting of the FN sequence GRGDSP with an additional cystein at the C-terminus was used as affinity matrix, four major proteins were recovered in the EDTA eluate. The apparent M_r of these proteins in SDS-PAGE were 145/130/115/90 kDa under nonreducing conditions and 160/130/110/100 kDa after reduction (Fig. 1A). To identify the bands, the GRGDSPC-binding proteins were analyzed by immunoblotting. As shown in Fig. 1B, the bands with a M_r of 130 and 90 kDa (nonreduced) reacted with an antiserum against $\alpha v\beta 3$. The 90-kDa band was also recognized by

an antiserum specific for the β3-subunit. The 115-kDa component was recognized by an antiserum against the β1-subunit. Based on the known apparent M_r in SDS-PAGE of β1, αv, and β3, they would correspond to the reduced bands of 130, 110, and 100 kDa, respectively. However, none of these or other tested antisera reacted with the 145-kDa band. Thus, the GRGDSPC-binding fraction contained the integrin αvβ3 and the integrin β1-subunit associated with an unidentified α-subunit (145 kDa nonreduced, 160 kDa reduced).

The two heterodimers were found to be separated when applied to a GRGDS-Sepharose. As shown in Fig. 2, the integrin αvβ3 was retained on the GRGDS-Sepharose (lanes 3, 4), while all of the unidentified β1-integrin passed through the column and could subsequently be recovered essentially free from αvβ3 by use of the GRGDSPC-Sepharose (lanes 5, 6). For comparison, α5β1 eluted from Sepharose conjugated with a 105-kDa FN fragment (lanes 1, 2) and α1β1 isolated on collagen-Sepharose (lanes 7, 8) are shown. These two integrins did not bind to any of the peptide columns, as expected, since α5β1 requires larger FN fragments for stable binding, and α1β1 recognizes yet unidentified structures which differ from RGD. Conversely, the peptide-binding integrins did not bind to Sepharose conjugated with collagen or to the 105-kDa FN fragment, which contains a GRGDSPA sequence. A 70/75-kDa protein (nonreduced/reduced) was obtained in variable amounts from the GRGDS column (visible in lane 4). The protein was not obtained from purified hepatocytes, but was recovered when blood was subjected to the purification procedure (not shown). This component was not further studied.

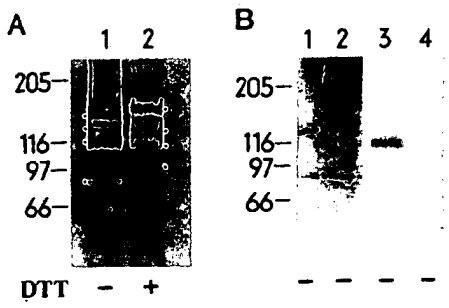


FIG. 1. Affinity chromatography of liver extract on GRGDSPC-Sepharose. (A) WGA-binding proteins isolated from rat liver as described under Materials and Methods were applied to a column of Sepharose conjugated with GRGDSPC peptide. After washing of the column, proteins were eluted with buffer containing EDTA, subjected to SDS-PAGE in unreduced (lane 1) or reduced form (lane 2), and stained with silver. The migration of size marker proteins and their apparent M_r (kDa) are indicated. The four major protein bands of the samples are marked with dots. (B) The proteins eluted from GRGDSPC-Sepharose (shown in lane 1) were analyzed in reduced form by immunoblotting as described under Materials and Methods. The antisera used were anti-αvβ3 (lane 1), anti-β3 (lane 2), anti-β1 (lane 3), and preimmune serum (lane 4). The migration of the size marker proteins is indicated.

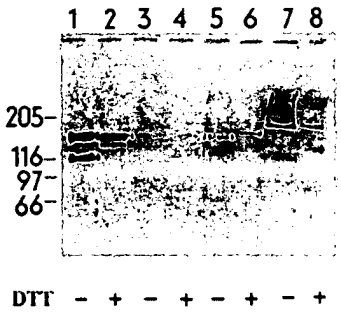


FIG. 2. SDS-PAGE of integrins isolated from rat liver. WGA-binding proteins from rat liver were applied sequentially to four different columns. EDTA-eluted proteins from the Sepharose columns conjugated with a 105-kDa FN fragment (lanes 1 and 2), GRGDS peptide (lanes 3 and 4), GRGDSPC peptide (lanes 5 and 6), and collagen type I (lanes 7 and 8) were analyzed by SDS-PAGE in unreduced (lanes 1, 3, 5, 7) and reduced (lanes 2, 4, 6, 8) forms. The migration of size marker proteins and their apparent M_r (kDa) are indicated.

In order to identify the α-subunit, it was isolated from 50 livers by the procedure shown in Fig. 2 followed by preparative SDS-PAGE to remove the β1-subunit. After transfer to PVDF membrane the N-terminal amino acid sequence of the protein was determined. A search in the Swiss Protein Data base revealed that the 20-amino-acid-long sequence obtained was related to the N-terminal parts of integrin α-subunits (Fig. 3A). Among the subunits known to associate with β1, the number of identical amino acids were the highest for α5 (60%) and the lowest for α1 and α2 (35%). The relatively low degree of homology makes it unlikely that α9 is the rat counterpart to any of the known human integrin subunits. Furthermore, a polyclonal antiserum raised against rat α5 did not recognize the 145/160-kDa subunit (Fig. 5). An internal amino acid sequence of the protein (Fig. 3B) also showed some homology to human integrin α-subunits, but was clearly distinct. Since the protein has unique amino acid sequences at two different positions, it is unlikely to be a splice variant of a previously described integrin α-subunit. According to the present nomenclature it is referred to as α9 in the following.

Amino acid sequencing of the integrin subunit recognized by antibodies against β3, and of an internal peptide generated from the integrin subunit reacting with antibodies against β1, confirmed that they are the rat homologues to these human proteins (Fig. 4).

Immunization with a synthetic peptide ($\alpha 9_{1-16}$) made from the N-terminal sequence of α9 (amino acids 1–16) resulted in an antiserum which reacted with α9 in immunoprecipitation (data not shown) and in immunoblotting as shown in Fig. 5. It recognized both the unreduced and the reduced forms of α9. After long exposure,

A

α9(r)	Y	N	I	D	A	Q	R	P	V	R	F	Q	-	-	G	P	P	G	S	F	F	G
α1(r)	F	*	V	*	V	K	N	S	M	S	*	S	-	-	*	*	V	E	D	M	*	*
α2(h)	*	*	V	G	L	P	E	A	K	I	*	S	-	-	*	*	S	S	E	Q	*	*
α3(h)	F	*	L	*	T	R	F	L	*	V	K	E	-	A	*	N	*	*	*	L	*	*
α4(h)	*	*	V	*	T	E	S	A	L	L	Y	*	-	-	*	*	H	N	T	L	*	*
α5(h)	F	*	L	*	*	E	A	*	A	V	L	S	-	-	*	*	*	*	*	*	*	*
α6(h)	F	*	L	*	T	R	E	D	N	V	I	R	K	Y	*	D	*	*	*	L	*	*
α7(h)	F	*	L	*	V	M	G	A	L	*	K	E	-	-	*	E	*	*	*	L	*	*
α8(h)	F	*	L	*	E	E	K	L	T	V	Y	S	-	-	*	*	*	*	*	Y	*	*
αv(h)	F	*	L	*	V	D	S	*	A	E	Y	S	-	-	*	*	E	*	*	Y	*	*

B

α9(r)	Q	V	-	-	-	-	P	F	E	Q	N	-	C	T	X	E	D	-	C	A	A	D	L	Q	L	R	
α1(r)	H	I	-	-	-	-	*	*	A	K	D	-	*	G	N	K	E	R	*	I	S	*	*	N	*	N	
α2(h)	S	I	-	-	-	-	*	*	A	K	D	-	*	G	E	G	*	L	*	I	S	*	*	V	*	D	
α3(h)	E	*	-	-	-	-	Q	*	Q	K	E	-	*	G	P	D	N	K	*	E	S	N	*	*	M	*	
α4(h)	I	M	K	K	T	I	N	*	A	R	F	-	*	A	H	*	N	-	*	S	*	*	*	*	V	S	
α5(h)	K	A	-	-	-	-	Q	I	L	L	D	-	*	G	E	D	N	I	*	V	P	*	*	*	*	E	
α6(h)	D	*	H	-	-	-	H	*	L	K	E	G	*	G	D	D	N	V	*	N	S	N	*	K	*	E	
α7(h)	E	I	-	-	-	-	H	*	L	K	Q	G	*	G	D	D	K	I	*	Q	S	N	*	*	*	V	
α8(h)	*	A	-	-	-	-	Y	I	L	V	D	-	*	G	E	D	N	L	*	I	P	*	*	H	*	S	
αv(h)	*	A	-	-	-	-	H	I	L	L	D	-	*	G	E	D	N	V	*	K	P	K	*	E	V	S	(639)

X represents C or S in the α9 sequence

FIG. 3. Comparison of amino acid sequences of β1-associated integrin α-subunits. The N-terminal amino acid sequence (A) and an internal amino acid sequence (B) of the integrin α9-subunit from rat is shown at the top. Below are the rat subunit α1 and the human subunits α2–α8 and αv with amino acids identical to α9 marked with an asterisk. Gaps (-) are inserted to optimize homologies; (h) and (r) denote human and rat sequences, respectively. Information on the human amino-terminal sequences was from Kramer *et al.* 1991 [37] for α2–α7 and αv and from Bossy *et al.* 1991 [51] for α8. The other human sequences were from Takada *et al.* 1991 [52] for α2–α6, α8, and αv and from Song *et al.* 1992 [53] for α7. The rat α1 sequence was from Ignatius *et al.* 1990 [34]. The location of the αv peptide in the protein is indicated by the number of the C-terminal serine.

a faint reactivity with α5 could also be detected, possibly due to a stretch of four identical amino acids in the α5 and the $\alpha 9_{1-16}$ peptides (amino acids 13–16, see Fig. 3A). No cross-reactivity with α1, αv, β1, or β3 was observed (not shown).

Purification of Integrins from Different Organs

To determine if the integrin α9β1 was expressed in other tissues than liver, the same fractionation procedure was also applied to spleen, kidney, heart muscle, and skeletal muscle. From all organs tested, an integrin was eluted from the GRGDSPC column which migrated identically to α9 in SDS–PAGE and which reacted with the anti-α9 antiserum in immunoblotting (not shown). Estimation of the yields of integrins α5β1 and α9β1 (Table 1) shows that approximately equal amounts of α9β1 were isolated from all organs except liver, from which significantly more α9β1 was obtained. The integrin α5β1 was found to be even more abundant in the liver relative to the other organs. We also isolated α1β1 and αvβ3 from these organs; α1β1 was obtained in the largest amount from skeletal muscle, while αvβ3 was obtained in small but detectable amounts from all organs (data not shown). Although this method of quantification may not accurately reflect the actual amount of the integrins in the different organs, due to the semiquantitative nature of the technique and different recoveries of membrane proteins from different organs, it serves to illustrate that α9β1 has a widespread distribution.

	499																			519
β1(h)	H	C	E	C	S	T	D	E	V	N	S	E	D	M	D	A	Y	C	R	K
β1(r)	H	C	E	C	S	T	D	E	V	N	S	E	D	M	D	A	Y	C	R	K

	1														14
β3(h)	G	P	N	I	C	T	T	R	G	V	S	S	C	Q	
β3(r)	E	S	N	I	C	T	T	R	G	V	K	S	C	Q	

FIG. 4. Comparison of the rat β1 and β3 amino acid sequences with their human counterparts. The proteins from rat liver which copurified with α9 and αv, respectively, were sequenced as described under Materials and Methods and compared to their human counterparts; (h) and (r) denote the human sequence and rat sequence, respectively. Information on the human sequences was from Moyle *et al.* 1991 [54].

Peptide Specificity

Different variants of the GRGDSPC peptide were synthesized, coupled to Sepharose, and used in affinity

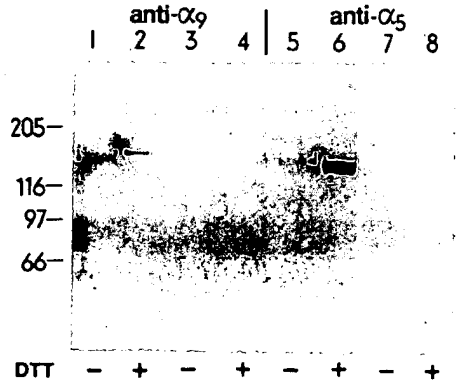


FIG. 5. Immunoblotting with anti-α9 antibodies. Integrins α5β1 (lanes 3, 4, 5, 6) and α9β1 (lanes 1, 2, 7, 8) were run on SDS–PAGE in unreduced form (lanes 1, 3, 5, 7) and in reduced form (lanes 2, 4, 6, 8) and transferred to nitrocellulose filter. The filter was cut into halves which were incubated with antisera against the α9-subunit (lanes 1–4) and the α5-subunit (lanes 5–8), respectively. The immunoreactive proteins were visualized by ECL. The migration of size marker proteins and their apparent M_r (kDa) are indicated.

chromatography of liver extracts. As shown in Fig. 6, a GRGESPC peptide was almost as efficient as the GRGDSPC peptide in the binding of $\alpha 9\beta 1$. The GKGDSPC and GKGESPC peptides, where the arginine residue of the two peptides above was substituted with lysine, had lower but still detectable binding activity. The cysteine, however, was critical; a GRGDSP peptide did not bind $\alpha 9\beta 1$ at all. The identity of the material eluted from the peptide–Sepharoses as $\alpha 9\beta 1$ was confirmed by immunoblotting with the anti-$\alpha 9$ antiserum (not shown).

No further proteins were released from the GRGDSPC–Sepharose by reducing agent (1 mg/ml DTT) after the EDTA elution, as monitored by silver staining of SDS–polyacrylamide gels (not shown). This result indicates that the binding of $\alpha 9\beta 1$ is not mediated by another protein bound to the thiol group of the peptide.

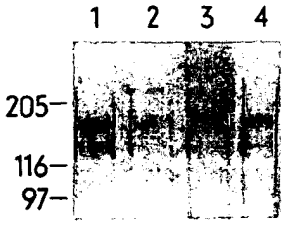


FIG. 6. Specificity of peptide binding of $\alpha 9\beta 1$. Variants of the GRGDSPC peptide were synthesized, coupled to Sepharose, and used in affinity chromatography of liver extracts. Equal volumes of the eluate from GRGDSPC–, GKGDSPC–, GKGESPC–, and GRGESPC–Sepharose were applied (in reduced form) to lanes 1, 2, 3, and 4, respectively, of an SDS–polyacrylamide gel. After electrophoresis, the proteins were stained with silver. The migration of size marker proteins and their apparent M_r (kDa) are indicated.

Ligand Specificity

By the use of a solid-phase binding assay, the ability of different ECM proteins to function as ligands for the integrin $\alpha 9\beta 1$ was tested. EHS–LN and collagen type I adsorbed to plastic microtiter wells were found to bind ^{125}I-labeled $\alpha 9\beta 1$ in a concentration-dependent manner, whereas the level of binding of $\alpha 9\beta 1$ to FN and fibrinogen was close to the background binding to BSA. VN bound some $\alpha 9\beta 1$ when coated at low concentrations, but at increasing ligand density less binding of receptor was obtained (Fig. 7A). A 50-fold excess of unlabeled $\alpha 9\beta 1$ reduced the binding of ^{125}I-labeled $\alpha 9\beta 1$ to LN and collagen by more than 60%, while the GRGDSPC peptide at concentrations up to 1 mg/ml did not affect the binding significantly (data not shown). To identify regions in EHS–LN and collagen I that interact with $\alpha 9\beta 1$, different fragments from the proteins were used in the binding assay (Fig. 7B). The integrin was found to bind to a fragment corresponding to the long arm of LN (E8), but not to the central cross domain of LN (P1). Of the three tested CNBr fragments of collagen $\alpha 1$(I), the CB8 fragment specifically bound the integrin, while the CB3 and the CB7 fragments were inactive. A triple-helical structure of CB8 and collagen was apparently required for the interaction with $\alpha 9\beta 1$, since termal denaturation at 50°C for 10 min prior to coating to the plastic wells at 37°C reduced the binding by >80 and 70%, respectively (not shown). These LN and collagen fragments were all active in promoting attachment of hepatocytes (Fig. 8), demonstrating that receptor-binding structures were also available in the plastic adsorbed fragments which did not bind $\alpha 9\beta 1$.

TABLE 1

Purification of Integrins from Different Organs

	Kidney	Heart	Liver	Spleen	Muscle
$\alpha 9\beta 1$	+	+	++	+	+
$\alpha 5\beta 1$	+	+++	++++	+	+

Note. Equal amounts of the indicated organs (12 g) were processed for isolation of integrins as described under Materials and Methods. The yield of purified proteins was measured by densitometric scanning of the $\beta 1$ band after SDS-PAGE and silver staining. For this purpose, an Ultroscan XL enhancer laser densitometer with the 2400 Gelscan software version 2.0 (Pharmacia) was used. The amount of $\alpha 5\beta 1$ obtained from liver was set to 100% (++++), and the symbols +++, ++, and + denote ~40, ~10, and 2–6% of this amount, respectively.

DISCUSSION

In this paper we describe the purification and initial characterization of an integrin from rat liver containing the $\beta 1$-subunit in association with a previously unidentified α-subunit.

The novel integrin subunit, which migrates similarly to $\alpha 2$ in SDS–PAGE, was observed in the EDTA eluate from GRGDSPC–Sepharose during affinity chromatography of solubilized rat liver. Since no integrin of the $\beta 1$-group except $\alpha v\beta 1$ [33] has been shown to bind short RGD-containing peptides in affinity chromatography, the α- and β-subunits were subjected to amino acid sequencing. A peptide sequence obtained for the β-subunit was identical to the human $\beta 1$, and the two 20-amino-acid-long sequences obtained from the α-subunit showed similarity with other integrin α-subunits, but were clearly distinct from previously known proteins. A comparison of the peptide sequences with the known $\beta 1$-associated α-subunits showed that the number of identical amino acids was in the same range as that obtained when the corresponding regions of the different human α-subunits are compared (35–60% for the N-terminus). The new subunit, called $\alpha 9$, is not likely a rat counterpart to any of the human α-subunits since the

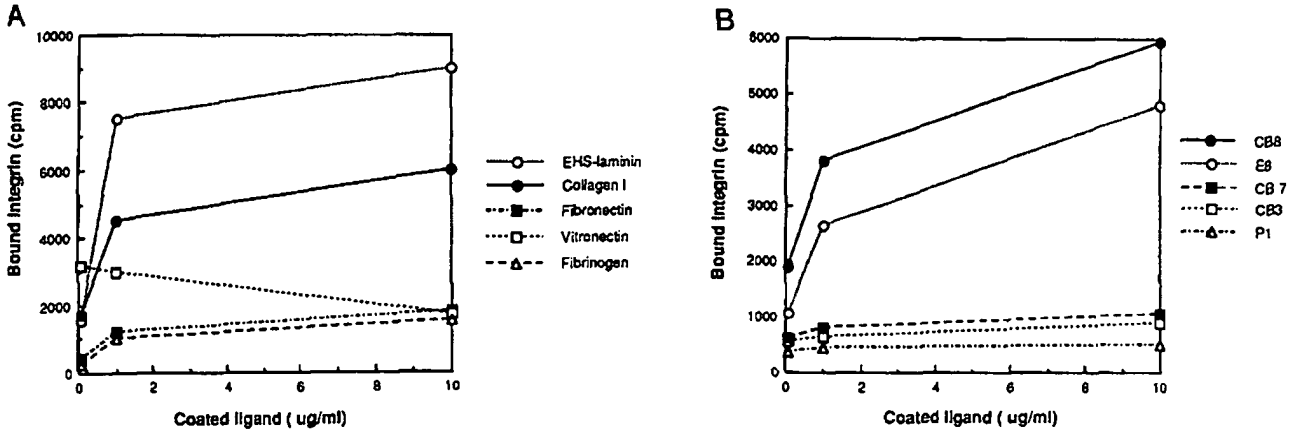


FIG. 7. Solid-phase receptor assay. ^{125}I-Labeled $\alpha 9\beta 1$ (10,000 cpm/ng) was incubated in microtiter wells (100,000 cpm/well) as described under Materials and Methods. The wells were previously coated with (A) EHS-LN, collagen type I, plasma FN, fibrinogen, VN or (B) fragments of LN (E8 and P1) and collagen I (CB3, CB7, and CB8), respectively, at the indicated concentrations. After washing, bound radioactivity was quantified in a gamma counter. Background binding to BSA-blocked wells was 250–300 cpm. The lowest coating concentration tested was 0.1 μg/ml.

degree of sequence similarity between homologous integrin subunits from different species is higher. For example, the identity between the rat and human $\alpha 1$-sequences is 80% [34], between the human and the chicken αv-subunits 80% [35, 36], and between the human and the mouse $\alpha 7$ it is 92% [37]. Furthermore, the highest similarity of $\alpha 9$ (N-terminus) was to $\alpha 5$, which can be distinguished from $\alpha 9$ based on migration in SDS–PAGE (Fig. 2) and antibody reactivity (Fig. 5). The $\alpha 9$-subunit belongs to the group of integrin subunits which do not have a light chain, as reduction resulted in slower migration in SDS–PAGE and no release of a low-molecular-weight fragment.

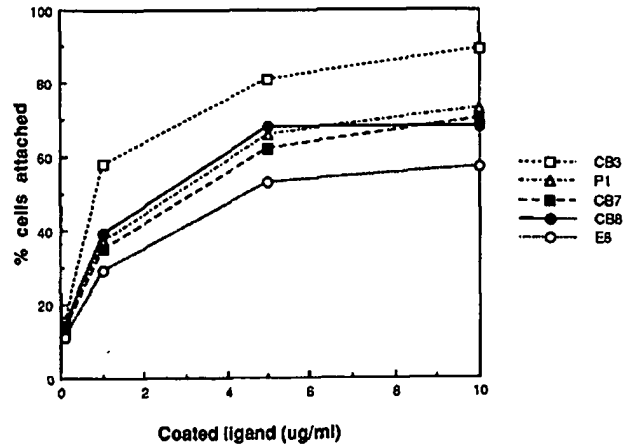


FIG. 8. Attachment of hepatocytes to LN and collagen fragments. Hepatocytes (4×10^5 cells) were seeded in wells (10 mm in diameter) coated with the indicated concentrations of the fragments P1, E8 (from EHS-LN), CB3, CB7, and CB8 (from collagen $\alpha 1$(I)), respectively. The lowest coating concentration tested was 0.1 μg/ml. After incubation at 37°C for 60 min, the dishes were washed [17] and the number of attached cells was determined by the hexosaminidase assay [55]. The number of cells attached to dishes coated with 10 μg/ml of intact collagen I was set to 100%.

The binding specificity was analyzed with different peptides similar to the GRGDSPC peptide, coupled to Sepharose. The RGD sequence was not absolutely required for binding of $\alpha 9\beta 1$; peptides in which the aspartic acid was changed to a glutamic acid bound $\alpha 9\beta 1$ almost as well as the GRGDSPC peptide. Substitution of arginine with a lysine resulted in reduced binding activity of the peptides (GKGDSPC and GKGESPC), but $\alpha 9\beta 1$ in low amounts could still be recovered also from these affinity matrices. However, the cysteine was critical for the interaction since a GRGDSP peptide lacked binding activity. The cysteine-containing peptides coupled to the columns existed as monomers, as shown by the mass spectrometry analysis. Recently the integrin $\alpha 2\beta 1$ was reported to interact with cyclic, but not with linear, RGD-containing peptides [8]. The authors suggested that a rigid conformation of the peptide was needed in order to bind $\alpha 2\beta 1$. Even if a very small amount of the peptides on the column we used occurred as dimers formed by a disulfide bridge between two peptides coupled to the column, the interaction between $\alpha 9\beta 1$ and the peptides is not dependent on a rigid conformation since soluble peptides could elute the integrin from the column (data not shown).

The specificity of the integrin $\alpha 9\beta 1$ for potential physiological ligands was investigated using a solid-phase receptor assay. As the integrin is widespread and could be purified from all organs analyzed, we tested extracellular proteins with a widespread distribution. FN, VN,

and fibrinogen, proteins known to bind integrins by the use of RGD sequences, were found to bind $\alpha 9\beta 1$ poorly. Surprisingly, ^{125}I-labeled $\alpha 9\beta 1$ bound specifically to EHS-LN and collagen I in this assay. The regions of EHS-LN and collagen $\alpha 1$(I) that interacted with the integrin were located on the E8 fragment of LN and on the CB8 fragment of collagen $\alpha 1$(I). In contrast, $\alpha 9\beta 1$ was not retained during affinity chromatography on Sepharose conjugated with EHS-LN or collagen I, illustrating that results obtained by these techniques have to be interpreted with caution.

Two of the fragments which did not bind $\alpha 9\beta 1$, CB7 and P1, contain an RGD sequence, while both of the $\alpha 9\beta 1$-binding fragments lack this motif. Further, the binding of ^{125}I-labeled $\alpha 9\beta 1$ to LN- or collagen-coated plastic could not be inhibited by soluble RGD-containing peptides, indicating that other or additional mechanisms are involved in the interactions. Collagen- and LN-binding integrins have in most cases been found not to recognize RGD sequences. However, RGD-dependent binding of integrins to native collagen has been described [8, 18], and $\alpha v\beta 3$ and $\alpha 5\beta 1$ are probably able to interact with RGD sequences in collagen if the collagen molecule is denatured [19, 38]. Similarly, RGD-dependent binding of cells to LN has been shown [16, 39, 40], although there are indications that the RGD site in LN is cryptic in the native molecule [41]. $\alpha v\beta 3$, which usually binds its ligands via RGD sequences, has been implicated as a LN receptor [16], but in one study the interaction of this integrin with LN was concluded to be RGD-independent [21].

The ligand specificity of $\alpha 9\beta 1$ was also surprising in view of the ability of antibodies against the integrin $\alpha 1$-subunit to completely block hepatocyte attachment to collagen type I [19]. These results indicate that $\alpha 1\beta 1$ is a dominant receptor on hepatocytes during formation of initial contacts with collagen type I. The same $\alpha 1$ antibody only marginally inhibits the binding of hepatocytes to LN in the same type of experiment (Forsberg and Johansson, unpublished observations), suggesting that LN receptors other than $\alpha 1\beta 1$ are active on these cells. The integrin $\alpha 9\beta 1$ is a possible candidate for such a function. To investigate if $\alpha 9\beta 1$ was present on hepatocytes, rat livers were perfused and the different cell types were separated as previously described [17]. $\alpha 9\beta 1$ could be isolated both from hepatocytes and from the nonparenchymal fraction-containing endothelial cells, Ito cells, and Kupffer cells (data not shown).

To search for natural ligands to this integrin we also looked for the presence of sequences similar to GRGDSPC in extracellular matrix proteins. Human collagen type I contains two RGD sequences in the $\alpha 1$(I) chain and three RGD sequences in the $\alpha 2$(I) chain [42, 43]. At least two of these are not conserved between species and none of them has a cysteine close to the RGD sequence. Alternatively, the binding site could be built up from parts of more than one collagen chain in the triple-helical molecule. The only RGD sequence in mouse EHS-LN is located in the P1 region, and it is not followed by a cysteine. Somewhat related sequences are present in S-LN (GDAPC) [44] and in the B2 chain of human LN (RGCTPC) [45]. Examples of other proteins with similar sequences are thrombin (RGDAC) [46], fibulin (RGGGPC) [47], thrombospondin (RGDAC) [48], and nidogen (RGDGQTC) [49]. The two latter proteins both have a number of repeats containing RGD-like sequences followed by cysteines [49, 50].

In summary, we have purified a new member of the integrin family, $\alpha 9\beta 1$, which is present in several organs. $\alpha 9\beta 1$ is efficiently purified on Sepharose conjugated with a cysteine-containing RGD peptide. EHS-LN and collagen type I are two candidates as physiological ligands for this receptor based on solid-phase binding data. Since neither of these proteins bind $\alpha 9\beta 1$ in an RGD-dependent manner additional ligands are likely to exist.[2]

This work was supported by grants from the Swedish Medical Research council (7147), the Swedish Cancer Society, King Gustaf V's 80th Anniversary Fund, the Medical Faculty at Uppsala University, the Odd Fellow orden, and the Swedish Society for Medical Research.

REFERENCES

1. Hemler, M. E. (1990) *Annu. Rev. Immunol.* **8**, 365–400.
2. Humphries, M. J. (1990) *J. Cell Sci.* **97**, 585–592.
3. Hynes, R. O. (1992) *Cell* **69**, 11–25.
4. Pierschbacher, M. D., and Ruoslahti, E. (1984) *Nature* **309**, 30–33.
5. Pierschbacher, M. D., and Ruoslahti, E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5985–5988.
6. Taub, R., Gould, R. J., Garsky, V. M., Ciccarone, T. M., Hoxie, J., Friedman, P. A., and Shattil, S. J. (1989) *J. Biol. Chem.* **264**, 259–265.
7. Yamada, K. M., and Kennedy, D. W. (1987) *J. Cell Physiol.* **130**, 21–28.
8. Cardarelli, P. M., Yamagata, S., Taguchi, I., Gorcsan, F., Chiang, S. L., and Lobl, T. (1992) *J. Biol. Chem.* **267**, 23159–23164.
9. Pierschbacher, M. D., and Ruoslahti, E. (1987) *J. Biol. Chem.* **262**, 7294–7298.
10. D'Souza, S. E., Ginsburg, M. H., and Plow, E. F. (1991) *Trends Biosci.* **16**, 246–250.
11. Obara, M., Kang, M. S., and Yamada, K. M. (1988) *Cell* **53**, 649–657.
12. Dufour, S., Duband, J-L., Humphries, M., Obara, M., Yamada, K. M., and Thiery, J. P. (1988) *EMBO J.* **7**, 2661–2671.
13. Edwards, J. G., Hameed, H., and Campbell, G. (1988) *J. Cell Sci.* **89**, 507–513.

[2] When this paper was ready for submission the sequence of the human integrin subunit $\alpha 9$ was reported (Palmer *et al.* (1993) *J. Cell Biol.* **123**, 1289–1297). The partial amino acid sequences presented in this study are 75% identical to the human $\alpha 9$, which is why the protein most likely represents the rat counterpart of human $\alpha 9$.

14. Johansson, S., Forsberg, E., and Lundgren, B. (1987) *J. Biol. Chem.* **262,** 7819–7824.
15. Mould, A. P., and Humphries, M. J. (1991) *EMBO J.* **10,** 4089–4095.
16. Goodman, S. L., Aumailley, M., and von der Mark, H. (1991) *J. Cell Biol.* **113,** 931–941.
17. Forsberg, E., Paulsson, M., Timpl, R., and Johansson, S. (1990) *J. Biol. Chem.* **265,** 6376–6381.
18. Dedhar, S., Ruoslahti, E., and Pierschbacher, M. D. (1987) *J. Cell Biol.* **104,** 585–593.
19. Gullberg, D., Gehlsen, K. R., Turner, D. C., Åhlen, K., Zijenah, L. S., Barnes, M. J., and Rubin, K. (1992) *EMBO J.* **11,** 3865–3873.
20. Elices, M. J., Urry, L. A., and Hemler, M. H. (1991) *J. Cell Biol.* **12,** 169–181.
21. Kramer, R. H., Cheng, Y., and Clyman, R. (1990) *J. Cell Biol.* **111,** 1233–1243.
22. Smilenov, L., Forsberg, E., Zeligman, I., Sparrman, M., and Johansson, S. (1992) *FEBS Lett.* **302,** 227–230.
23. Woods, A, Couchman, J. R., Johansson, S., and Höök, M. (1986) *EMBO J.* **5,** 665–670.
24. Timpl, R., Paulsson, M., Dziadek, M., and Fujiwara, S. (1987) *Methods Enzymol.* **145,** 363–390.
25. Knorr, R., Trzeciak, A., Bannwarth, W., and Gillesen, D. (1989) *Tetrahedron Lett.* **30,** 1927–1930.
26. King, D. S., Fields, C. G., and Fields, G. B. (1990) *Int. J. Pept. Protein Res.* **36,** 255–266.
27. Bottger, B. A., Hedin, U., Johansson, S., and Thyberg, J. (1989) *Differentiation* **41,** 158–167.
28. Pujades, C., Forsberg, E., Enrich, C., and Johansson, S. (1992) *J. Cell Sci.* **102,** 815–820.
29. Harlow, E., and Lane, D. (1988) *in* Antibodies: A Laboratory Manual, pp. 82–83, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
30. Blobel, G., and Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
31. Burnette, W. N. (1981) *Anal. Biochem.* **112,** 195–203.
32. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038.
33. Vogel, B. E., Tarone, G., Giancotti, F. G., Gailit, J., and Ruoslahti, E. (1990) *J. Biol. Chem.* **265,** 5934–5937.
34. Ignatius, M. J., Large, T. H., Houde, M., Tawil, J. W., Barton, A., Esch, F., Carbonetto, S., and Reichardt, L. F. (1990) *J. Cell Biol.* **111,** 709–720.
35. Suzuki, S., Argrawes, W. S., Pytela, R., Arai, H., Krusius, T., Pierschbacher, M. D., and Ruoslahti, E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8614–8618.
36. Bossy, B., and Reichardt, L. F. (1990) *Biochemistry* **29,** 10191–10198.
37. Kramer, R. H., Vu, M. P., Cheng, Y. F., Ramos, D. M., Timpl, R., and Waleh, N. (1991) *Cell Regul.* **2,** 805–817.
38. Pfaff, M., Aumailley, M., Specks, U., Knolle, J., Zerwes, H. G., and Timpl, R. (1993) *Exp. Cell Res.* **206,** 167–176.
39. Tashiro, K., Sephel, G. C., Greatorex, D., Sasaki, M., Shirashi, N., Martin, G. R., Kleinman, H. K., and Yamada, Y. (1991) *J. Cell. Physiol.* **146,** 451–459.
40. Grant, D. S., Tashiro, K., Segui-Real, B., Yamada, Y., Martin, G. R., and Kleinman, H. (1989) *Cell* **58,** 933–943.
41. Aumailley, M., Gerl, M., Sonnenberg, A., Deutzmann, R., and Timpl, R. (1990) *FEBS Lett.* **262,** 82–86.
42. Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C. T., Jaenisch, R., and Prockop, D. (1988) *Biochem. J.* **253,** 919–922.
43. Pikkarinen, T., Kallunki, T., and Tryggvason, K. (1988) *J. Biol. Chem.* **263,** 6751–6758.
44. Hunter, D. D., Shah, V., Merlie, J. P., and Sanes, J. R. (1989) *Nature* **338,** 229–234.
45. Kuivaniemi, H., Tromp, G., Chu, M-L., and Prockop, D. (1988) *Biochem. J.* **252,** 633–640.
46. Furie, B., Bing, D. H., Feldmann, R. J., Robinson, D. J., Burnier, J. P., and Furie, B. C. (1982) *J. Biol. Chem.* **257,** 3875–3882.
47. Argraves, W. S., Tran, H., Burgess, W. H., and Dickerson, K. (1990) *J. Cell Biol.* **111,** 3155–3164.
48. Lawler, J., and Hynes, R. O. (1986) *J. Cell Biol.* **103,** 1635–1648.
49. Mann, K., Deutzmann, R., Aumailley, M., Timpl, R., Raimondi, L., Yamada, Y., Pan, T., Conway, D., and Chu, M. (1989) *EMBO J.* **8,** 65–72.
50. Sun, X., Skorstensgaard, K., and Mosher, D. F. (1992) *J. Cell Biol.* **118,** 693–701.
51. Bossy, B., Bossy-Wetzel, E., and Reichardt, L. F. (1991) *EMBO J.* **10,** 2375–2385.
52. Takada, Y., Murphy, E., Pil, P., Ginsburg, M. H., and Hemler, M. E. (1991) *J. Cell Biol.* **115,** 257–266.
53. Song, W. K., Wang, W., Foster, R., Bielser, D. A., and Kaufman, S. J. (1992) *J. Cell Biol.* **117,** 643–657.
54. Moyle, M., Napier, M. A., and Mclean, J. W. (1991) *J. Biol. Chem.* **266,** 19650–19658.
55. Landegren, U. (1984) *J. Immunol. Methods* **67,** 379–388.

Received February 4, 1994

Cell, Vol. 96, 319–328, February 5, 1999, Copyright ©1999 by Cell Press

The Integrin αvβ6 Binds and Activates Latent TGFβ1: A Mechanism for Regulating Pulmonary Inflammation and Fibrosis

John S. Munger,[5,8] Xiaozhu Huang,[1,8]
Hisaaki Kawakatsu,[1,8] Mark J. D. Griffiths,[1,8]
Stephen L. Dalton,[1,3] Jianfeng Wu,[1]
Jean-François Pittet,[1,2] Naftali Kaminski,[1]
Chrystelle Garat,[4] Michael A. Matthay,[2,4]
Daniel B. Rifkin,[5,6] and Dean Sheppard[1,4,7]
[1]Lung Biology Center
[2]Department of Anesthesia
[3]Department of Dermatology
[4]Cardiovascular Research Institute
and Department of Medicine
University of California, San Francisco
San Francisco, California 94143-0854
[5]Department of Medicine
[6]Cell Biology and Kaplan Cancer Center
New York University School of Medicine
New York, New York 10016-6402

Summary

Transforming growth factor β (TGFβ) family members are secreted in inactive complexes with a latency-associated peptide (LAP), a protein derived from the N-terminal region of the TGFβ gene product. Extracellular activation of these complexes is a critical but incompletely understood step in regulation of TGFβ function in vivo. We show that TGFβ1 LAP is a ligand for the integrin αvβ6 and that αvβ6-expressing cells induce spatially restricted activation of TGFβ1. This finding explains why mice lacking this integrin develop exaggerated inflammation and, as we show, are protected from pulmonary fibrosis. These data identify a novel mechanism for locally regulating TGFβ1 function in vivo by regulating expression of the αvβ6 integrin.

Introduction

The transforming growth factor β (TGFβ) family consists of three closely related isoforms (TGFβ1, -2, and -3) that are prototypes of the larger TGFβ superfamily. In vitro, TGFβs exert nearly identical effects that can be grouped into three broad areas: modulation of inflammatory cell function, growth inhibition and differentiation, and control of extracellular matrix production. Studies of animal models as well as human clinical specimens strongly suggest that TGFβs are important in the pathogenesis of several diseases, including fibrotic conditions (Broekelmann et al., 1991; Border et al., 1992; Sime et al., 1997). TGFβ1 knockout mice develop diffuse mononuclear cell infiltrates that prove lethal within a few weeks from birth (Shull et al., 1992; Kulkarni et al., 1993). In contrast, TGFβ2 and TGFβ3 knockout mice display only developmental defects (Kaartinen et al., 1995; Sanford et al., 1997). Major differences among TGFβ isoform functions in vivo are due at least in part to differences in the promoter regions of the various isoform genes (Taipale et al., 1998). It is also possible, but not proven, that there are TGFβ isoform-specific mechanisms for converting latent TGFβs to the active forms.

The TGFβs are secreted as complexes composed of three proteins derived from two genes. Each *TGFβ* gene encodes a procytokine consisting of a C-terminal TGFβ sequence and a larger N-terminal region that, after processing, forms a protein called latency-associated peptide (LAP). LAP and TGFβ remain noncovalently associated, and in this configuration TGFβ is unable to bind to its receptors; that is, TGFβ is latent. In most cases, the complex of LAP and TGFβ (the small latent complex SLC) is joined by latent TGFβ binding protein 1 (LTBP1), a matrix protein with sequence similarity to the fibrillins, and the complex of all three proteins is called the large latent complex (LLC). Latent TGFβ can be linked by LTBP to binding sites in the extracellular matrix (Taipale et al., 1996).

The mechanisms involved in activating latent TGFβ are not fully understood, but recently there has been important progress in this area. Plasmin can activate latent TGFβ in cell-free systems (Lyons et al., 1990) and in cell culture (Sato et al., 1990). However, plasminogen knockout mice display none of the pathologic features of TGFβ knockout mice, suggesting that plasmin is unlikely to be the only molecule activating TGFβ. Reactive oxygen species can activate TGFβ in vitro (Barcellos-Hoff and Dix, 1996), and radiation treatment appears able to activate TGFβ in vivo via this mechanism (Barcellos-Hoff et al., 1994). Thrombospondin (TSP) 1 can activate TGFβ by binding to a defined site on LAP and inducing a conformational change in the latent complex; TGFβ is then bound to TSP1 in an active state (Schultz-Cherry and Murphy-Ullrich, 1993; Schultz-Cherry et al., 1995). A recent study of the similar patterns of inflammation exhibited by TGFβ1 and TSP1 knockout mice suggests that TSP1 is a major activator of TGFβ1 in vivo (Crawford et al., 1998). However, the inflammatory changes in the TSP1 knockout mice are not nearly as severe as those in TGFβ1 knockout mice, suggesting overlapping mechanisms of TGFβ activation.

LAP-β1 and LAP-β3 contain arginine-glycine-aspartic acid (RGD) sequences, which are also binding site motifs in ligands for a subset of integrins. LAP-β1 can bind effectively to one such integrin, αvβ1, but the functional role of LAP–integrin interactions is not known (Munger et al., 1998). Integrins were first identified based on their roles in mediating cell attachment and migration but have recently been recognized to participate in more complex cellular events, including survival (Meredith et al., 1993), proliferation, and regulation of gene expression (Werb et al., 1989). The fact that αvβ1 can bind latent TGFβ suggests that this or other RGD-binding integrins might regulate TGFβ bioactivity.

The integrin αvβ6 is expressed principally on epithelial cells, where it has been shown to be a receptor for RGD sites in fibronectin (Weinacker et al., 1994), tenascin (Prieto et al., 1993), and vitronectin (Huang et al., 1998a).

[7]To whom correspondence should be addressed (e-mail: deans@itsa.ucsf.edu).
[8]These authors contributed equally to this work.

αvβ6 is expressed at low levels in healthy adult lung tissues but is rapidly upregulated by injury and inflammation (Breuss et al., 1995). Inactivation of the β6 subunit gene in mice revealed an unexpected role for αvβ6 in downregulating inflammatory responses to minor environmental insults in the lungs and skin (Huang et al., 1996). Somewhat surprisingly, despite exaggerated skin and lung inflammation, $\beta6^{-/-}$ mice do not develop fibrosis at either site. The combination of enhanced inflammation and protection from fibrosis suggested a localized deficiency of active TGFβ1 as a cause of the $\beta6^{-/-}$ phenotype. We therefore sought to determine whether TGFβ1 LAP is a ligand for αvβ6 and whether interaction of αvβ6 with LAP-containing complexes can lead to latent TGFβ1 activation. To determine whether such an effect might have relevance to disease, we also utilized $\beta6^{-/-}$ mice in a well-characterized model of pulmonary fibrosis induced by bleomycin, a model that has previously been shown to be critically dependent on TGFβ (Giri et al., 1993).

Results

TGFβ1 LAP Is a Ligand for the Integrin αvβ6

To determine whether LAP-β1 could bind αvβ6, we performed affinity chromatography by passing labeled secreted αvβ6 over Sepharose cross-linked to recombinant LAP, the known αvβ6 ligand fibronectin, or bovine serum albumin (BSA; to detect nonspecific binding). Bound proteins were eluted by EDTA, since interactions of integrins with ligands require the presence of divalent cations. Bands corresponding to truncated αv (130 kDa) and β6 (85 kDa) were eluted by EDTA from LAP- or fibronectin-Sepharose, but not from BSA-Sepharose (Figure 1A). The identity of the 85 kDa band as β6 was confirmed by Western blotting and by immunoprecipitation (Figures 1B and 1C). To demonstrate that full-length αvβ6 also binds to LAP, we repeated affinity chromatography with unlabeled octylglucoside lysates of β6-transfected SW480 cells (Figure 1D). A 95 kDa protein corresponding to full-length β6 was detected by Western blotting in eluted fractions from LAP-Sepharose but not from BSA-Sepharose.

To determine the effects of αvβ6/LAP interactions on cells, we performed cell adhesion assays with β6-transfected SW480 cells. LAP-coated wells supported αvβ6-dependent adhesion of β6-transfected cells, but mock-transfected cells did not adhere to any concentration of LAP (Figure 2B). Essentially identical results were obtained with mock- and β6-transfected 293 cells, Chinese hamster ovary (CHO) cells, and NIH 3T3 cells (not shown). β6-transfected SW480 cells, but not mock transfectants, also adhered to dishes coated with large latent TGFβ1 complexes (LLC; Figure 2C). Adhesion to LAP and LLC was abolished by anti-αvβ6 antibody 10D5 and was unaffected by antibodies against β1 (P5D2) or αvβ5 (P1F6; not shown). To determine whether αvβ6 mediated adhesion to LAP through an interaction with the RGD sequence, we performed cell adhesion assays with mutant LAP containing a D-to-E substitution mutation within the RGD site (Figure 2D). β6-transfected SW480 cells did not attach to any concentration of mutant LAP. Furthermore, adhesion of β6-transfected, but not mock-transfected, SW480 cells to either LAP or equimolar concentrations of either small or large latent TGFβ1 complexes containing LAP induced phosphorylation of two downstream integrin-signaling intermediates, the focal adhesion kinase (FAK) and paxillin (Figure 2E). Phosphorylation of each protein was completely inhibited by addition of the αvβ6-blocking antibody 10D5 (not shown).

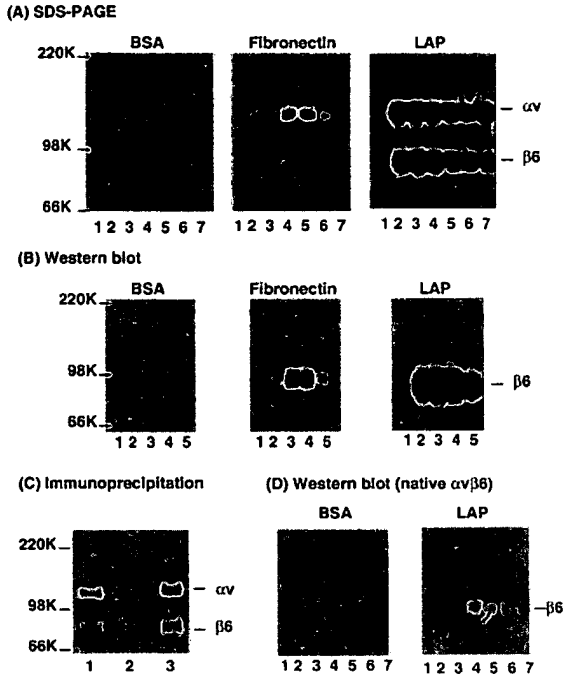


Figure 1. Affinity Chromatography

(A) ^{35}S-labeled secreted αvβ6 was incubated with either BSA-, fibronectin- or LAP-Sepharose. Bound proteins were eluted with EDTA and analyzed by SDS-PAGE under nonreducing conditions. Lane 1 was the final fraction eluted with column buffer, lanes 2–6 were eluted with EDTA, and lane 7 was eluted with 8 M urea.

(B) Western blot of eluted fractions with anti-β6 antibody 4B5. Lane 1 was the final fraction eluted with column buffer, lanes 2–4 were eluted with EDTA, and lane 5 was eluted with 8 M urea.

(C) Immunoprecipitation of CHO supernatant (lane 1) and CHO supernatant proteins eluted by EDTA from BSA (lane 2) and LAP (lane 3) columns with anti-αvβ6 MAb R6G9.

(D) Western blot of proteins from octylglucoside lysates of β6-transfected SW480 cells eluted from BSA- or LAP-Sepharose columns. Lane 1 was the final fraction eluted with column buffer, lanes 2–5 were eluted with EDTA, and lane 6 was eluted with 8 M urea. Molecular size markers (in kDa) are shown to the left.

β6-Transfected Cells Induce TGFβ Activity

To determine if binding to αvβ6 activates TGFβ1, we cocultured four different β6-expressing cells with mink lung epithelial reporter cells stably expressing a portion of the plasminogen activator inhibitor 1 promoter (TMLC) (Abe et al., 1994). For all four lines, coculture with β6-expressing cells caused a significant increase in luciferase levels compared to coculture with control cells (Figure 3A). These increases were abolished by MAbs against active TGFβ or αvβ6. A different reporter cell line (NIH 3T3 cells transfected with PAI1/luciferase) yielded

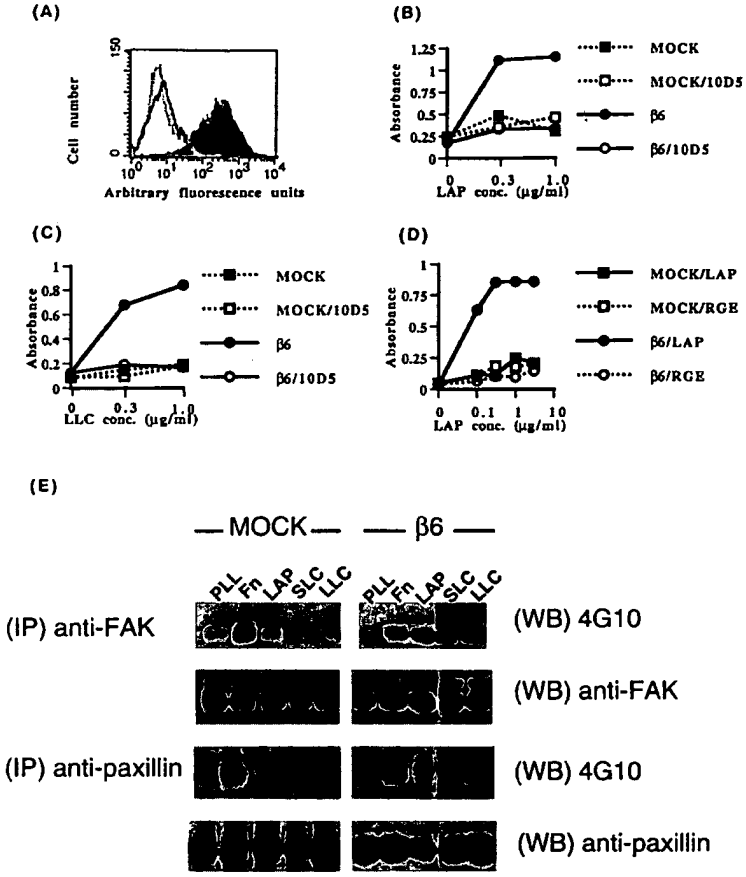


Figure 2. Adhesion of β6-Transfected Cells to LAP

(A) Control and β6-transfected SW480 cells were stained with anti-αvβ6 mAB E7P6 (white peaks represent control cells; black peaks represent β6 transfectants) and analyzed by flow cytometry. Dotted lines represent β6-transfected cells incubated with PBS.

(B and C) Nontransfected cells and β6-transfected cells were allowed to attach to wells coated with increasing concentrations of LAP (B) or LLC (C) or with 1% BSA. Prior to plating, cells were incubated with or without anti-αvβ6 antibody 10D5.

(D) Adhesion of SW480-β6 cells to recombinant LAP containing a single glutamic acid for aspartic acid substitution mutation in the RGD site (RGE LAP) was compared with adhesion to authentic recombinant LAP.

(E) Lysates of mock- and β6-transfected SW480 cells plated for 30 min on poly-L-lysine (PLL) or on poly-L-lysine plus fibronectin (Fn), LAP, SLC, or LLC were immunoprecipitated with antibodies against FAK or paxillin followed by Western blotting with either anti-phosphotyrosine antibody 4G10, anti-FAK, or anti-paxillin.

similar results (data not shown). Three cell lines that are capable of adhering to immobilized LAP via αvβ1 (293, MG63, and A549 cells) (Munger et al., 1998) did not activate TGFβ in similar assays (Figure 3A and unpublished data). Antibodies against the integrin β1 subunit or the integrin αvβ5 had no effect on activation (not shown).

To determine whether TGFβ activation by αvβ6 required cell–cell contact, we did coculture assays with inserts to separate reporter and β6-expressing cells by a few millimeters while allowing soluble molecules to pass. In the absence of contact, β6-expressing cells caused a slight induction of luciferase activity, but induction was minimal compared to β6-expressing cells in contact with the reporter cells (Figure 3B). These results indicate that the active TGFβ generated by αvβ6 is most efficiently detected by cells in contact with the β6-expressing cells, but that at least a small amount of the active TGFβ formed is freely diffusible.

To determine if increased secretion of latent TGFβ by β6-expressing lines could account for the results, serum-free medium conditioned by each line was tested for total TGFβ activity. β6-transfected CHO cells secreted more latent TGFβ (8-fold) than did mock-transfected CHO cells. However, in the other three cell types, latent TGFβ secretion was higher in the control lines (data not shown). All four lines secreted TGFβ1 as the predominant TGFβ isoform. Cocultures in serum-free conditions yielded results essentially identical to those presented, so latent TGFβ secreted by the cocultured cells is sufficient for measurable TGFβ activation. To determine whether the observed active TGFβ was specifically TGFβ1, we added isoform-specific neutralizing antibodies against TGFβ1, β2, and β3 to cocultures containing β6-transfected SW480 cells. Anti-TGFβ1 blocked luciferase induction, whereas anti-TGFβ2 and anti-TGFβ3 did not (Figure 3C). In addition, recombinant LAP (which both neutralizes TGFβ in solution and binds the αvβ6 integrin) blocked TGFβ activation.

αvβ6-Mediated Activation of TGFβ1 Does Not Require Other Known Activators of TGFβ

We next tested whether αvβ6-induced activation of TGFβ1 was occurring through previously described mechanisms of TGFβ activation. Activation of TGFβ by cocultures of endothelial cells and vascular smooth muscle requires plasmin (Sato et al., 1990), binding of mannose-6-phosphate on LAP (Dennis and Rifkin, 1991), and incorporation of LLC into the ECM via tissue transglutaminase (Kojima et al., 1993; Nunes et al., 1997). Therefore, we tested the effects of inhibitors that block each of these steps: the plasmin inhibitor aprotinin, M6P, inhibitors of transglutaminase-mediated cross-linking (cystamine and monodansylcadaverine), and a polyclonal antibody against the N terminus of LTBP1 (Ab450) that blocks LTBP linkage to the ECM (Nunes et al., 1997).

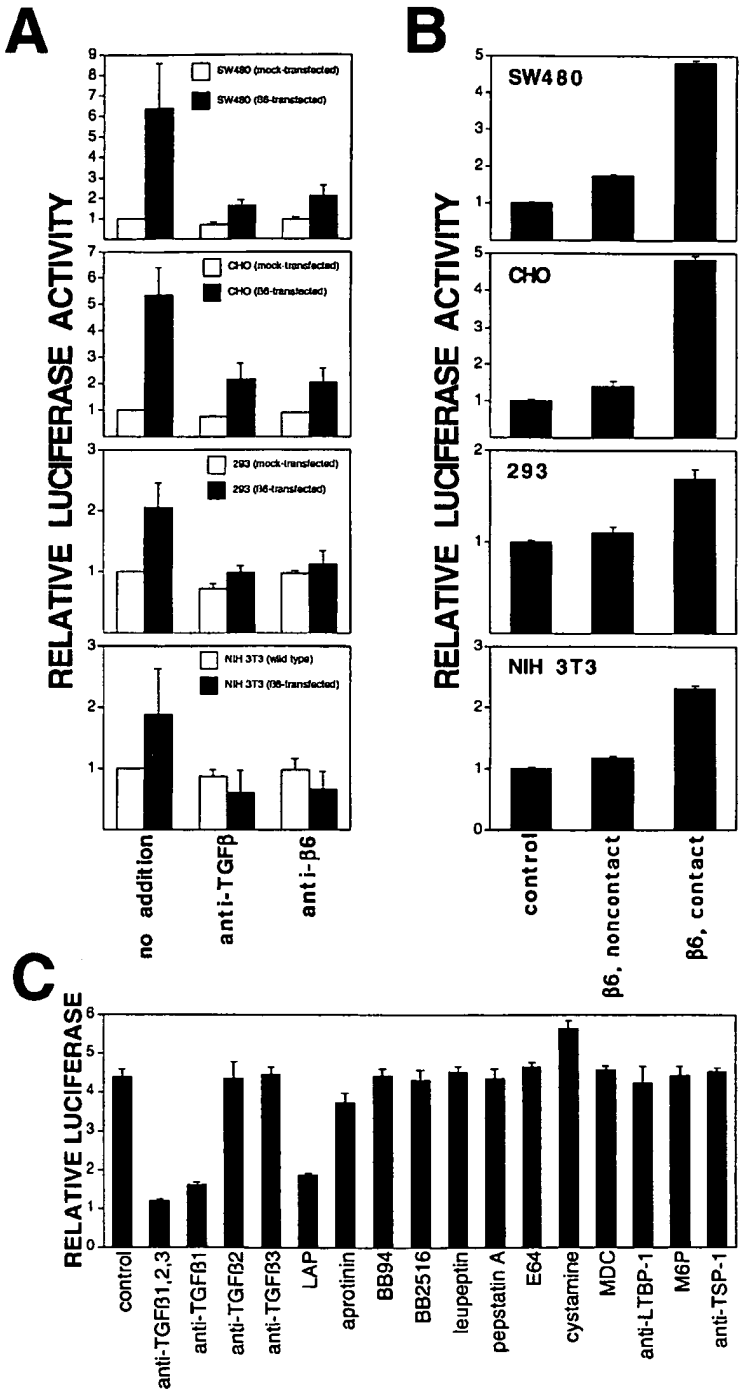


Figure 3. β6-Expressing Cells Activate TGFβ

(A) Equal numbers of reporter and test cells were cultured 16–20 hr and lysed for measurement of luciferase activity. Results are for four test cell lines. Additions are shown at the bottom (anti-TGFβ: MAb 1D11, 10 μg/ml; anti-β6: MAb 10D5, 10 μg/ml). Relative luciferase activity is the measured activity divided by the activity of the coculture with mock-transfected cells. Results are the mean (±SEM) of at least three experiments done in duplicate.

(B) The effect of close proximity between reporter and test cells on TGFβ activation was determined using culture inserts. Equal numbers of reporter cells in the bottom well, test cells in the bottom well, and test cells in the insert were cocultured for 16–20 hr and luciferase activity of the reporter cells was measured. Relative luciferase activity is measured activity divided by the activity of reporter cells cultured with control cells in both the bottom well and the insert. Results are the means (±SEM) of triplicate measurements.

(C) Activation of TGFβ by β6-expressing cells does not require the activity of proteases or molecules involved in other systems of TGFβ activation and involves only TGFβ isoform 1. Mock- or β6-transfected SW480 cells were cocultured with reporter cells as described in (3A). Additions are indicated at the bottom. Data represent the mean (±SEM) of quadruplicate measurements.

Because other proteases can activate TGFβ in vitro (Munger et al., 1998), we tested inhibitors of metallo-, aspartic, and cysteine proteases (BB94, BB2516, leupeptin, pepstatin A, and E64). Finally, TSP1-mediated activation of TGFβ1 can be blocked by MAb 133 (Schultz-Cherry et al., 1994). None of these inhibitors blocked the activation observed in cocultures with β6-expressing cells (Figure 3C).

Binding of LAP to αvβ6 Integrin Is Not Sufficient for Latent TGFβ1 Activation

To determine whether binding to αvβ6 is sufficient for activation of TGFβ or whether additional interactions with cell components are required, we examined the effects of truncation mutations of the β6 subunit cytoplasmic domain. Three mutants were examined; of these, only mutant 777T, which lacks the last 11 amino

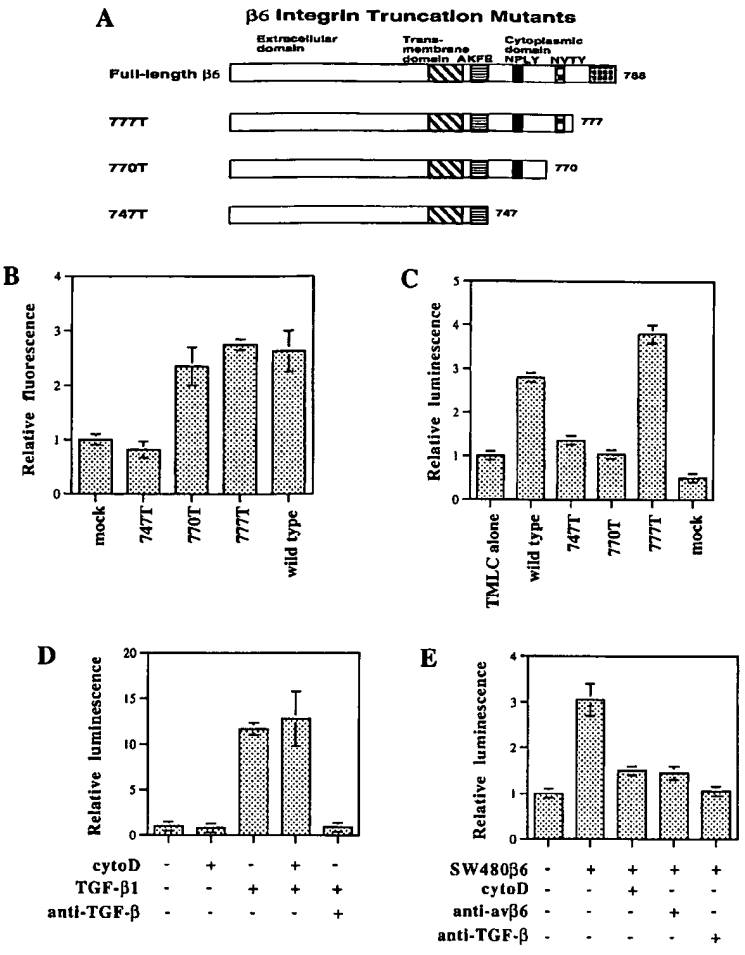


Figure 4. Binding of LAP to αvβ6 Integrin Is Not Sufficient for Latent TGFβ Activation

(A) Schematic representation of the wild-type β6 integrin subunit (full-length β6) and the three truncation mutants studied.

(B) Mock- or various β6-transfectants were incubated in suspension with 0.1 μg/ml LAP for 30 min at 37°C and analyzed by flow cytometry with anti-LAP MAb VB3A9. Relative fluorescence is the mean fluorescence of each sample divided by the mean fluorescence of mock transfectants. Results are the mean (±SEM) of at least three experiments.

(C) Equal numbers of reporter and test cells were cultured for 16–20 hr and luciferase activity was measured. For (C–E), relative luciferase activity is the measured activity divided by the activity of the TMLC alone and results are the mean (±SEM) of at least three experiments done in duplicate.

(D) TMLC were cultured for 16–20 hr in the presence or absence of cytochalasin D (100 μm), human recombinant TGFβ1 (10 pM), or anti-TGFβ (MAb 1D11, 10 μg/ml).

(E) Equal numbers of reporter cells and test cells were cultured for 16–20 hr in the presence or absence of cytochalasin D (100 μm), anti-αvβ6 integrin (MAb 1OD5, 50 μg/ml), or anti-TGFβ (MAb 1D11, 10 μg/ml).

acids of the β6 cytoplasmic domain, localizes to focal contacts (Cone et al., 1994). To determine whether deletions in the β6 cytoplasmic domain would affect latent TGFβ1 binding, transfectants were incubated with LAP and analyzed by flow cytometry with antibody to LAP. More LAP was detected on the surface of cells expressing wild-type αvβ6 than on mock transfectants (Figure 4B). Mutant 747T showed no binding above background, but both 770T and 777T showed LAP binding similar to wild-type αvβ6. In coculture assays, cells expressing mutants 747T and 770T showed little or no activation of latent TGFβ (Figure 4C). In contrast, mutant 777T activated latent TGFβ. No consistent difference was detected in total TGFβ secreted by these transfectants (data not shown). Since 770T bound LAP but failed to activate latent TGFβ, binding of LAP by αvβ6 is not sufficient for activation of latent TGFβ.

To determine whether an intact cytoskeleton is required for TGFβ activation, we cocultured reporter cells with β6-transfected SW480 cells in the presence of 100 μm cytochalasin D, conditions under which 100 percent of the cells became round but remained adherent. Cytochalasin D did not inhibit LAP binding to the cell surface, nor did it affect surface expression of αvβ6 or TGFβ secretion (not shown). Cytochalasin D had no effect on the ability of reporter cells to respond to active TGFβ added to the culture medium (Figure 4D). However, cytochalasin D added to cocultures of β6 transfectants and reporter cells blocked activation of latent TGFβ (Figure 4E). The inhibition was similar to that achieved by antibody to αvβ6 or to TGFβ. Thus, an intact cytoskeleton is required for αvβ6-mediated activation of TGFβ.

$\beta 6^{-/-}$ Mice Are Protected against Bleomycin-Induced Pulmonary Fibrosis

Pulmonary fibrosis was evaluated by examination of lung morphology and by measurement of hydroxyproline content in $\beta 6^{+/+}$ and $\beta 6^{-/-}$ 129 strain mice at 15, 30, and 60 days after intratracheal instillation of bleomycin. Fibrosis was significant in bleomycin-treated wild-type mice by 30 days and progressed to 60 days (Figure 5A), whereas in $\beta 6^{-/-}$ mice, lung morphology remained nearly unaltered throughout the experiment, with only small patches of fibrosis (Figure 5B), and the lung hydroxyproline content was not significantly different from that measured in saline-treated animals. Similar results were obtained in offspring of 129 by C57Bl/6 intercrosses (not shown). These results suggest that expression of αvβ6 is required for pulmonary fibrosis in response to bleomycin.

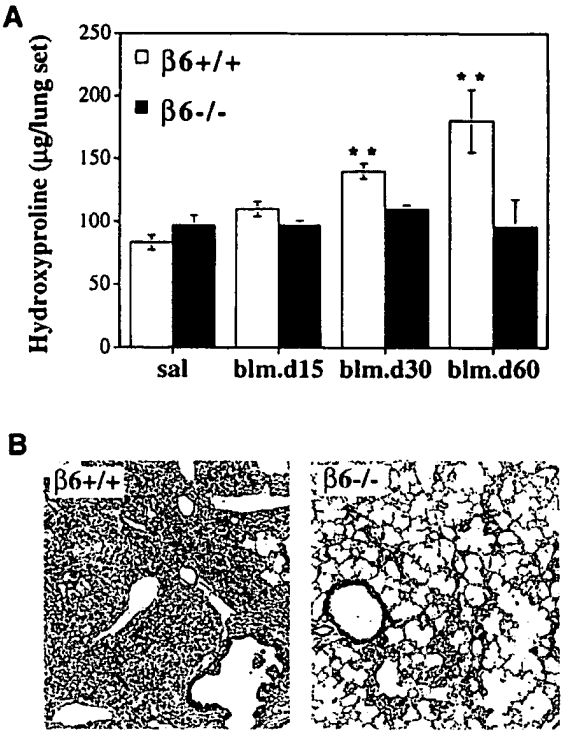


Figure 5. $\beta 6^{-/-}$ Mice Are Protected against Bleomycin-Induced Pulmonary Fibrosis

(A) Bleomycin (0.03 iu: blm) induces pulmonary fibrosis in $\beta 6^{+/+}$ but not $\beta 6^{-/-}$ mice, indicated by an elevation of lung hydroxyproline content compared with saline (sal)-treated controls 30 and 60 days after administration. Data from saline-treated mice at each time point are combined, as there was no significant difference between groups. Data are expressed as means (± SEM) of five to seven observations, ** $p < 0.01$.

(B) Histology of low power sections (magnification 200×) demonstrates dense accumulation of collagenous extracellular matrix in lungs of bleomycin-treated $\beta 6^{+/+}$ but not $\beta 6^{-/-}$ mice 60 days after injection.

To determine whether the resistance of $\beta 6^{-/-}$ mice to bleomycin-induced lung injury and fibrosis was due to a blunted inflammatory response, we counted inflammatory cells obtained from bronchoalveolar lavage (BAL) or minced lungs from $\beta 6^{-/-}$ and $\beta 6^{+/+}$ mice after treatment with saline and 5 and 15 days after treatment with bleomycin. Bleomycin increased the total cell counts and the numbers of neutrophils, lymphocytes, and macrophages in both lines of mice, but the effects were always greater in $\beta 6^{-/-}$ mice. These findings are consistent with our previous report of enhanced lung inflammation in $\beta 6^{-/-}$ mice and suggest that protection from bleomycin-induced pulmonary fibrosis is not due to inhibition of the inflammatory response to bleomycin.

To determine whether the exaggerated inflammation and protection from injury and fibrosis in $\beta 6^{-/-}$ mice was due to impaired synthesis of TGFβ1, we analyzed TGFβ protein expression by immunohistochemistry and by the TMLC bioassay on eluates from lung slices that had been heated to 80°C for 20 min to release and activate TGFβ. Specificity of the bioassay was confirmed by >80% inhibition of all samples by anti-TGFβ1 antibody. TGFβ eluted from lung slices was not different between lines or between saline and bleomycin treatment (relative luciferase activity compared to TMLC alone): saline, 6.5 ± 1.6 (mean ± SEM) for $\beta 6^{+/+}$ mice and 6.3 ± 1.4 for $\beta 6^{-/-}$ mice; bleomycin, 5.5 ± 1.6 for $\beta 6^{+/+}$ mice and 5.2 ± 1.1 for $\beta 6^{-/-}$ mice. Furthermore, immunohistochemistry with an antibody against a 30-amino-acid C-terminal peptide of TGFβ1 (LC1-30) under conditions reported to detect both active and inactive TGFβ (Barcellos-Hoff et al., 1995) revealed the presence of TGFβ throughout the lungs and airways in both saline- and bleomycin-treated animals, with no detectable increase at any time point after bleomycin treatment in either line of mice. Immunohistochemistry under conditions reported to detect only active TGFβ demonstrated little staining at any time point (data not shown).

αvβ6 Protein Expression Is Focally Induced by Bleomycin in $\beta 6^{+/+}$ Mice

We have previously reported that αvβ6 is expressed at low levels in skin and lung epithelium, but that expression is dramatically upregulated in cutaneous wounds and in injured and inflamed epithelia (Breuss et al., 1995). To determine whether bleomycin treatment produced similar increases in αvβ6 expression, we performed immunohistochemistry on lung sections from $\beta 6^{+/+}$ mice 10 days after treatment with either saline or bleomycin. As expected, no αvβ6 immunoreactivity was seen in lungs from $\beta 6^{-/-}$ mice. Diffuse, low-level expression of αvβ6 was apparent in airway and alveolar epithelial cells in $\beta 6^{+/+}$ mice treated with saline, whereas focal areas with markedly increased expression of αvβ6 were present throughout the lungs of bleomycin-treated animals (Figure 6A).

Keratinocytes and Airway Epithelial Cells Activate TGFβ1 through αvβ6

To determine directly whether αvβ6 expressed in mouse skin could activate TGFβ1, we performed bioassays by coculturing keratinocytes obtained from $\beta 6^{-/-}$ or $\beta 6^{+/+}$ mice with TMLC (Figure 6C). $\beta 6^{+/+}$ cells expressed abundant amounts of αvβ6 (Figure 6B) and demonstrated TGFβ1 activity in this assay (Figure 6C), whereas $\beta 6^{-/-}$ cells did not induce TGFβ1 activity. Because of the difficulty of culturing murine lung epithelial cells, we performed similar studies using primary cultures of human bronchial epithelial cells, which also demonstrated significant expression of αvβ6 (Figure 6B). These cells also induced αvβ6-dependent TGFβ1 activity.

Discussion

In this report, we show that LAP-β1 is a ligand for the integrin αvβ6 and that αvβ6-expressing cell lines can activate endogenous latent TGFβ1. Furthermore, $\beta 6^{-/-}$ mice are protected from bleomycin-induced pulmonary fibrosis, a model that has been shown to be critically dependent on TGFβ activity. In the mice we studied, TGFβ1 was constitutively expressed in the lungs, and the amount of total TGFβ protein was not demonstrably different in $\beta 6^{-/-}$ and $\beta 6^{+/+}$ mice and was not significantly affected by treatment with bleomycin. However,

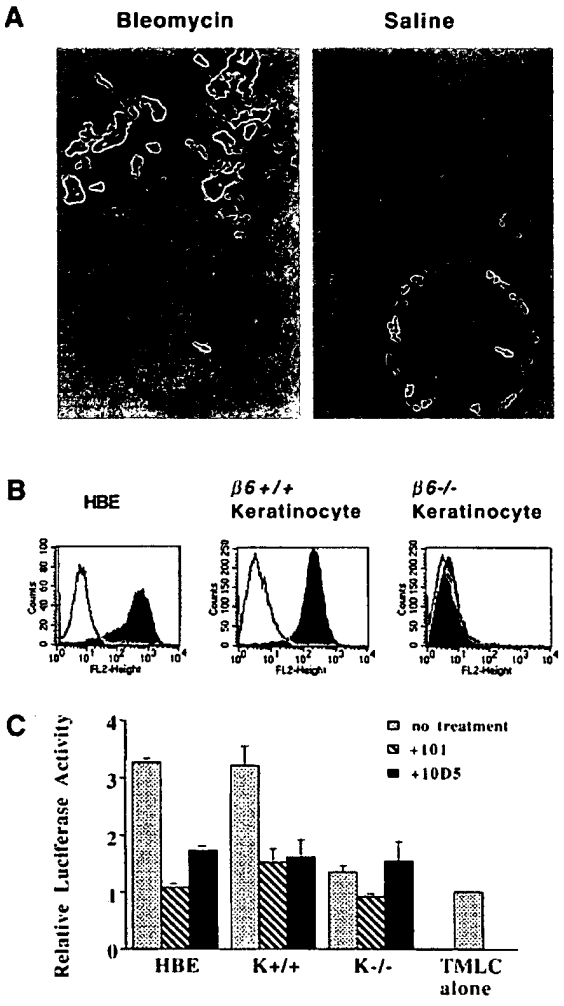


Figure 6. Bleomycin Focally Increases αvβ6 Expression in $\beta6^{+/+}$ Mice, and Keratinocytes and Airway Epithelial Cells Expressing αvβ6 Activate TGFβ1

(A) Representative sections from lungs of $\beta6^{+/+}$ mice 10 days after treatment with saline or bleomycin. Arrows, normal staining of conducting airway epithelium; arrowheads, alveolar epithelial cells with dramatically increased αvβ6 expression.

(B) Primary cultures of $\beta6^{+/+}$ and $\beta6^{-/-}$ keratinocytes or human bronchial epithelial cells (HBE) were stained with anti-αvβ6 antibodies 10D5 or E7P6 (shaded histograms) or PBS and analyzed by flow cytometry.

(C) $\beta6^{+/-}$ and $\beta6^{-/-}$ keratinocytes or HBE were cocultured with TMLC for 16–20 hr in the presence or absence of anti-αvβ6 antibody 10D5 or anti-TGFβ1 antibody (101). Relative luciferase activity is measured activity divided by activity of TMLC alone. Data are the mean (±SEM) of at least four measurements.

in response to bleomycin, αvβ6 expression was dramatically increased in the lungs of $\beta6^{+/+}$ mice. Together with the observations that $\beta6^{-/-}$ mice develop inflammation in the skin and lungs (partially reproducing the TGFβ1 knockout phenotype), the results of this study indicate that the regulated expression of αvβ6 by epithelia is important for local activation of TGFβ1 in response to injury and inflammation. This idea is consistent with a model in which tissue injury induces αvβ6 expression, which in turn locally activates TGFβ1 already abundantly present in many tissues. TGFβ1, once activated, enhances matrix deposition (healing or fibrosis) and downregulates the inflammatory response to injury.

This feedback model highlights the fact that resolution of inflammation is an active process. There are two regulatory pathways that might allow rapid amplification of this antiinflammatory feedback system. First, TGFβ itself induces β6 integrin subunit expression (Sheppard et al., 1992; Wang et al., 1996). Second, TGFβ1 induces its own expression (Van Obberghen-Schilling et al., 1988). Presumably, mechanisms exist to reverse these positive feedback effects; these mechanisms may fail in pathologic states of persistent TGFβ activity and fibrosis that involve epithelia.

The β6 knockout mice develop inflammation only in skin and lung and not in other tissues where β6 is expressed (e.g., uterus, renal epithelium, urinary bladder). This selectivity could be a consequence of the unique susceptibility of the skin and lung to environmental insults, leading to subclinical inflammation that must be actively repressed. For example, skin involvement occurs in areas most exposed to physical trauma, and lung inflammation in $\beta6^{-/-}$ mice is worse when mice are housed in unventilated cages. However, mice expressing a null mutation in the TGFβ1 gene develop exaggerated inflammation in multiple organs (Shull et al., 1992). In addition, in keeping with the known effects of TGFβ1 in inhibiting proliferation of epithelial cells, these mice demonstrate increased mitoses and epithelial hyperplasia in multiple epithelial organs. Despite careful morphologic examination of the liver, pancreas, bladder, stomach, uterus, and intestine, we have been unable to identify any of these abnormalities in $\beta6^{-/-}$ mice. Together, these data demonstrate that binding to αvβ6 is not the principal mechanism of TGFβ1 activation in most organs and that the developmental effects of TGFβ1 do not require activation by interaction with this integrin. Whereas other activation mechanisms are involved in developmental effects of TGFβ1, interaction with αvβ6 appears to be important for locally titrating the augmented TGFβ1 activity required in response to injury, at least in the lungs and skin.

Previous studies of TGFβ activation suggested the critical involvement of proteases, particularly plasmin. Our results, along with other recent work, suggest that nonproteolytic mechanisms are important physiologic pathways leading to active TGFβ. TSP1-mediated activation occurs when TSP1 binds LAP-β1 at a site near its N terminus. This presumably induces a conformational change that activates the complex, although the active TGFβ1 molecule remains bound to TSP1. αvβ6 binds LAP at the RGD site located near the C terminus. The nonproteolytic mechanisms must rely on an intrinsic ability of LAP to adopt different conformations. Conformational flexibility of LAP has already been documented in circular dichroism studies that showed recombinant free LAP undergoing a major conformational change upon binding TGFβ in solution (McMahon et al., 1996).

While αvβ6 expression is clearly necessary for the TGFβ activation mechanism we describe, is it sufficient? If αvβ6 is not sufficient, its role might simply be to concentrate latent TGFβ at the cell surface, thereby permitting some separate mechanism to activate TGFβ. In

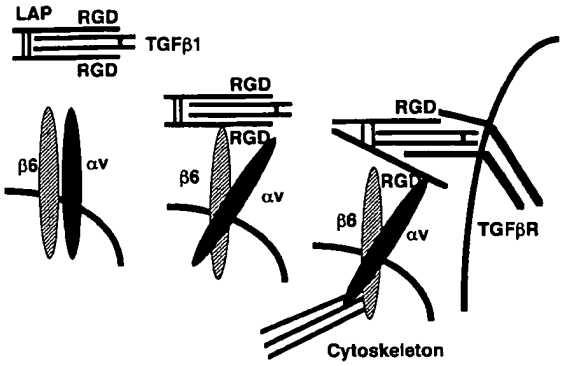


Figure 7. Model of TGFβ1 Activation by αvβ6

The data presented suggest that when latent TGFβ1 complexes bind to αvβ6, sites in the β6 cytoplasmic domain become accessible for binding to the actin cytoskeleton. Cytoskeleton-associated integrin then induces a change in the conformation of the latent complex, allowing access of mature TGFβ1 to TGFβ receptors and induction of classic TGFβ signaling.

attempting to answer this question, we tested whether molecules or processes known to be involved in other systems of TGFβ activation are required. Previous studies of endothelial and smooth muscle cell cocultures, which activate latent TGFβ, suggested that plasmin, the IGF-II/M6PR, and transglutaminase-mediated cross-linking of latent TGFβ to the ECM are all required for activation. However, our results indicate that none of these molecules or processes, nor TSP1 or a wide range of proteases, is involved in αvβ6-mediated activation of TGFβ. In addition, the fact that we observe activation using six different αvβ6-expressing cell lines and two reporter cell lines suggests that any additional molecules that are required must be widely expressed.

We have identified one additional requirement for activation: the ability of αvβ6 to connect with the actin cytoskeleton. Cells expressing mutant β6 were able to activate TGFβ only when the mutant integrin could localize to focal contacts, a process that involves clustering and mechanical linkage of integrins to the actin cytoskeleton in complexes containing an array of adapter proteins that includes FAK and paxillin. Cells expressing β6 mutants that do not localize to focal contacts do not activate latent TGFβ, even though one of these β6 mutants (770T) is still able to bind LAP via αvβ6. Cytochalasin D, which disrupts actin filaments, blocked TGFβ activation by cells expressing αvβ6. These results suggest that binding of latent TGFβ to αvβ6 per se is not sufficient for activation to occur; following binding, αvβ6 must also associate with the actin cytoskeleton in order to activate bound latent TGFβ (see model, Figure 7). Thus, modulation of cytoskeleton/αvβ6 interactions might be a means to regulate TGFβ1 activation independent of changes in αvβ6 expression.

Of the integrins known to bind RGD sequences, three are now known to bind to LAP-β1 (αvβ1, αvβ6, and, weakly, αvβ5), and one, the platelet integrin αIIbβ3, may (Grainger et al., 1995). The main functions heretofore ascribed to LAP are TGFβ latency and the facilitation of TGFβ secretion (Gray and Mason, 1990). The finding that multiple integrins can bind TGFβ1-LAP raises the possibility of an additional function, the ability to initiate signaling via integrins. The results of the present study demonstrate that LAP-containing latent TGFβ1 complexes can induce phosphorylation of at least two components of integrin-signaling complexes, FAK and paxillin. This finding raises the possibility that these "latent" complexes could initiate integrin-mediated effects on cell behavior.

The observation that αvβ6 induces TGFβ1 activity also suggests an alternative mechanism by which at least one integrin can affect cell behavior—by activating extracellular TGFβ1 that, in turn, initiates responses by binding to its own cognate receptor(s). This mechanism appears to explain the exaggerated lung and skin inflammation and protection from pulmonary fibrosis in $\beta 6^{-/-}$ mice and suggests the possibility of regulating local inflammation and fibrosis by targeting this integrin.

Experimental Procedures

Cell Lines, Antibodies, and Reagents

Cell lines were obtained from American Type Culture Collection and transfected with integrin expression plasmids as described (Weinacker et al., 1994). Mink lung epithelial cells stably transfected with a plasmid containing the luciferase cDNA downstream of a TGFβ-sensitive portion of the plasminogen activator inhibitor 1 promoter (TMLC) were used as described (Abe et al., 1994). Mouse anti-αvβ6 MAbs E7P6, R6G9 (Weinacker et al., 1994), and 10D5 (Huang et al., 1998a), rabbit anti-β6 MAbs 4B5 and B1 (Huang et al., 1998b), and mouse MAb VB3A9 against TGFβ1 LAP (Munger et al., 1998) were produced as described. Mouse anti-phosphotyrosine MAb 4G10 was obtained from Upstate Biotechnology, Inc. (Lake Placid, NY); mouse MAbs against FAK and paxillin were obtained from Transduction Laboratories (Lexington, KY); and rabbit polyclonal antibodies against FAK were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). LAP and LAP (RGE) were produced in a baculovirus system as described (Munger et al., 1998). Recombinant SLC and LLC were gifts of Drs. H. Ohashi and H. Tsumura (Kirin Brewery Co., Gunma, Japan). MAb 1D11 against active TGFβ (all isoforms), anti-TGFβ1 polyclonal chicken Ig (AF-101-NA), anti-TGFβ2 polyclonal goat IgG (AB-112-NA), and anti-TGFβ3 polyclonal goat IgG (AB-244-NA) were from R and D Systems, Minneapolis, MN. Anti-TSP1 MAb 133 (Schultz-Cherry and Murphy-Ullrich, 1993) was a gift of Dr. Murphy-Ullrich (University of Alabama, Birmingham). Rabbit polyclonal antiserum LC1-30 against a C-terminal peptide of TGFβ1 was a gift of Kathy Flanders (National Cancer Institute, Bethesda, MD). Anti-LTBP1 polyclonal rabbit antibody 450 was produced as described (Nunes, et al., 1997). Other reagents were all analytical grade.

Cell lines were grown in Dulbecco's modified Eagle's medium (DMEM) with 4.5 g/l glucose, L-glutamine, and 10% fetal bovine serum. Murine keratinocytes were obtained and grown as previously described (Huang et al., 1996). Human bronchial epithelial cells were purchased from Clonetics, grown in serum-free bronchial epithelial cell medium (Clonetics), and used at passage 1.

Affinity Chromatography

LAP, BSA, and chymotrypsin-digested fibronectin were coupled to cyanogen bromide–activated Sepharose essentially as described (Pytela et al., 1985). Affinity matrices contained 2.5 mg/ml of fibronectin, 4 mg/ml of BSA, and 7 mg/ml of LAP. Columns were washed and blocked with 1% BSA. Secreted αvβ6 was produced as described (Weinacker et al., 1994). Culture medium was passed through affinity columns, and bound proteins were eluted with column buffer, then with 20 mM EDTA in 50 mM Tris and 150 mM NaCl, and finally with 8 M urea. Octylglucoside lysates of β6-transfected SW480 cells were used for affinity chromatography under the same conditions with addition of 25 mM octylglucoside.

Immunoprecipitation

Samples were incubated with antibodies for 3 hr at 4°C. Immune complexes were collected by incubation for 1.5 hr with protein G-Sepharose. Beads were washed three times, boiled for 3 min in Laemli sample buffer, and then analyzed by SDS-PAGE and autoradiography.

Western Blotting

Proteins were separated by SDS-PAGE, transferred to a nylon membrane, and blocked for 1 hr in Tris-buffered saline containing 3% BSA or 5% skim milk. After incubation with primary antibody for 3 hr and then with peroxidase-conjugated secondary antibody for 1 hr, blots were developed with ECL (Amersham).

Cell Adhesion Assays

The assays were performed as previously described (Busk et al., 1992). Untreated polystyrene 96-well flat-bottom microtiter plates (Flow Laboratories, McLean, VA) were coated with LAP or 1% BSA. Cells were plated at 50,000 cells/well, and plates were centrifuged (top side up) at 10 g for 5 min and then incubated for 1 hr at 37°C. Nonadherent cells were removed by centrifugation, and attached cells were fixed, stained, and lysed with 50 μl of 2% Triton and quantified by measuring absorbance at 595 nm.

Flow Cytometry

Cells were blocked with normal goat serum, washed with PBS, and incubated with primary antibody for 20 min and then with phycoerythrin-conjugated secondary antibody (Boehringer Mannheim) for 20 min at 4°C. Cells were resuspended with PBS and analyzed by FACScan (Becton Dickinson, Rutherford, NJ).

TGFβ Bioassay

TMLC and test cells were suspended at 5×10^5 cells/ml in DMEM containing 10% FCS. TMLC were plated first at 50 μl per microtiter well (Microtest III plates, Falcon, Franklin Lakes, NJ) and allowed to attach for 1 hr. Keratinocytes and bronchial epithelial cells were suspended at 4-fold higher density. Medium was replaced with 50 μl/well of the same medium with or without additions (e.g., antibodies). Fifty microliters of test cell suspension or test solutions was added and plates were cultured for 16–20 hr. Lysates were assayed for luciferase activity as described (Abe et al., 1994). Similar cocultures were done in 24-well plates (Costar model 3526, Corning, NY) with inserts designed for attachment-dependent cell culture (Millicell-PCF 3 μm filter, Millipore, Bedford, MA), but 300 μl of reporter and test cells were added to upper and/or lower chambers. To elute TGFβ from lung slices, tissues were quick frozen in liquid nitrogen, and five 20 μm cryosections were incubated for 20 min in 500 μl of DMEM at 80°C.

Bleomycin Treatment

Age- and sex-matched 8- to 12-week-old $\beta6^{+/+}$ and $\beta6^{-/-}$ mice of strains 129/terSVEMS and 129/terSVEMS by C57Bl/6 were maintained in a specific pathogen-free environment. Bleomycin (Mead Johnson, Princeton, NJ) was dissolved in sterile saline (0.03 or 0.05 units in 60 μl). Bleomycin or saline was administered transtracheally under methoxyflurane anesthesia by direct cut down.

Hydroxyproline Assay

Hydroxyproline content was measured in whole mouse lungs by methods previously described with modifications (Woessner, 1961). Following perfusion with PBS and homogenization, samples were incubated on ice in tricarboxylic acid (50%; Sigma Chemical Co., St. Louis, MO) and baked in 12 N hydrochloric acid (Mallinckrodt Baker Inc., Paris, KY) for 24 hr at 110°C. Aliquots reconstituted with distilled water were added to 1.4% chloramine T (Sigma) in 10% isopropanol and 0.5 M sodium acetate for 20 min. Erlich's solution (Sigma) was added and incubated at 65°C for 15 min. Absorbance was measured at 550 nm.

Histology and Immunohistochemistry

The trachea and both lungs were fixed by inflation at 25 cm H_20 with 10% formalin and embedded in paraffin (for histology and total TGFβ staining with antibody LC1-30) or inflated with 50% OCT and quick frozen in liquid nitrogen. Five micrometer sections were stained with hematoxylin and eosin and with trichrome to identify extracellular collagen. Sections were fixed in cold acetone for β6 antibody (B1) or in methanol/acetone for "active" TGFβ staining with antibody (LC1-30). Sections were blocked with Peroxoblock (Zymed Lab) and Avidin/Biotin Blocking Kit (Vector) and rinsed and incubated with 3% goat serum in PBS for 15 min and then overnight at 4°C in primary antibody. Sections were incubated in biotin-labeled secondary antibody for 1 hr and in ABC avidin/peroxidase reagent (Vector Lab) for 1 hr at room temperature, and chromagen was developed using the DAB Plus Kit (Zymed).

Acknowledgments

This work was supported by NIH grants HL47412, HL53949, HL56385 (D. S.), HL51854, HL51586 (M. A. M.), and RO1CA23753 (D. B. R.) and General Clinical Research Center grants #M01 RR 00096, NCRR, and 5 K12 CA01713 (J. S. M.). J. S. M. acknowledges support from the Inger-Ma Sonneborn Fund. M. J. D. G. was supported by the Wellcome Trust UK. We thank Steve Rosen and Robert Pytela for thoughtful comments and acknowledge the editorial assistance of K. A. Hubbard.

Received October 30, 1998; revised December 9, 1998.

References

Abe, M., Harpel, J.G., Metz, C.N., Nunes, I., Loskutoff, D.J., and Rifkin, D.B. (1994). An assay for transforming growth factor-beta using cells transfected with a plasminogen activator inhibitor-1 promoter-luciferase construct. Anal. Biochem. *216*, 276–284.

Barcellos-Hoff, M.H., and Dix, T.A. (1996). Redox-mediated activation of latent transforming growth factor-beta 1. Mol. Endocrinol. *10*, 1077–1083.

Barcellos-Hoff, M.H., Derynck, R., Tsang, M.L., and Weatherbee, J.A. (1994). Transforming growth factor-beta activation in irradiated murine mammary gland. J. Clin. Invest. *93*, 892–899.

Barcellos-Hoff, M.H., Ehrhart, E.J., Kalia, M., Jirtle, R., Flanders, K., and Tsang, M.L. (1995). Immunohistochemical detection of active transforming growth factor-beta in situ using engineered tissue. Am. J. Pathol. *147*, 1228–1237.

Border, W.A., Noble, N.A., Yamamoto, Y., Harper, J.R., Yamaguchi, Y., Pierschbacher, M.D., and Ruoslahti, E. (1992). Natural inhibitor of transforming growth factor-beta protects against scarring in experimental kidney disease. Nature *360*, 361–364.

Breuss, J.M., Gallo, J., DeLisser, H.M., Klimanskaya, I.V., Folkesson, H.G., Pittet, J.F., Nishimura, S.L., Aldape, K., Landers, D.V., Carpenter, W., et al. (1995). Expression of the β6 integrin in development, neoplasia, and tissue repair suggests a role in epithelial remodeling. J. Cell Sci. *108*, 2241–2251.

Broekelmann, T., Limper, A.H., Colby, T.V., and McDonald, J.A. (1991). Transforming growth factor-beta 1 is present at sites of extracellular matrix gene expression in human pulmonary fibrosis. Proc. Natl. Acad. Sci. USA *88*, 6642–6646.

Busk, M., Pytela, R., and Sheppard, D. (1992). Characterization of the integrin αvβ6 as a fibronectin-binding protein. J. Biol. Chem. *267*, 5790–5796.

Cone, R.I., Weinacker, A., Chen, A., and Sheppard, D. (1994). Effects of beta subunit cytoplasmic domain deletions on the recruitment of the integrin alpha v beta 6 to focal contacts. Cell Adhes. Commun. *2*, 101–113.

Crawford, S.E., Stellmach, V., Murphy-Ullrich, J.E., Ribeiro, S.M., Lawler, J., Hynes, R.O., Boivin, G.P., and Bouck, N. (1998). Thrombospondin-1 is a major activator of TGF-beta1 in vivo. Cell *93*, 1159–1170.

Dennis, P.A., and Rifkin, D.B. (1991). Cellular activation of latent transforming growth factor beta requires binding to the cation-independent mannose-6-phosphate/insulin-like growth factor type II receptor. Proc. Natl. Acad. Sci. USA *88*, 580–584.

Giri, S., Hyde, D.M., and Hollinger, M.A. (1993). Effect of antibody

to transforming growth factor-beta on bleomycin induced accumulation of lung collagen in mice. Thorax *48*, 959–966.

Grainger, D.J., Wakefield, L., Bethell, H.W., Farndale R.W., and Metcalfe, J.C. (1995). Release and activation of platelet latent TGF-beta in blood clots during dissolution with plasmin. Nat. Med. *1*, 932–937.

Gray, A., and Mason, A. (1990). Requirement for activin A and transforming growth factor-beta 1 pro-region in homodimer assembly. Science *247*, 1328–1330.

Huang, X.Z., Wu, J.F., Cass, D., Erle, D.J., Corry, D., Young, S.G., Farese, R.V.J., and Sheppard, D. (1996). Inactivation of the integrin beta 6 subunit gene reveals a role of epithelial integrins in regulating inflammation in the lung and skin. J. Cell Biol. *133*, 921–928.

Huang, X.Z., Wu, J.F., Spong, S., and Sheppard, D. (1998a). The integrin alphavbeta6 is critical for keratinocyte migration on both its known ligand, fibronectin, and on vitronectin. J. Cell Sci. *111*, 2189–2195.

Huang, X.Z., Wu, J.F., Zhu, W., Pytela, R., and Sheppard, D. (1998b). Expression of the human integrin beta6 subunit in alveolar type II cells and bronchiolar epithelial cells reverses lung inflammation in beta6 knockout mice. Am. J. Respir. Cell. Mol. Biol. *19*, 636–642.

Kaartinen, V., Voncken, J.W., Shuler, C., Warburton, D., Bu, D. Heisterkamp, N., and Groffen, J. (1995). Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction. Nat. Genet. *11*, 415–421.

Kojima, S., Nara, K., and Rifkin, D.B. (1993). Requirement for transglutaminase in the activation of latent transforming growth factor-beta in bovine endothelial cells. J. Cell Biol. *121*, 439–448.

Kulkarni, A.B., Huh, C.G., Becker, D., Geiser, A., Lyght, M., Flanders, K.C., Roberts, A.B., Sporn, M.B., Ward, J.M., and Larlsson, S. (1993). Transforming growth factor-beta 1 null mutation in mice causes excessive inflammatory response and early death. Proc. Natl. Acad. Sci. USA *90*, 770–774.

Lyons, R.M., Gentry, L.E., Purchio, A.F., and Moses, H.L. (1990). Mechanism of activation of latent recombinant transforming growth factor-beta 1 by plasmin. J. Cell Biol. *110*, 1361–1367.

McMahon, G.A., Dignam, J.D., and Gentry, L.E. (1996). Structural characterization of the latent complex between transforming growth factor beta 1 and beta 1-latency-associated peptide. Biochem. J. *313*, 343–351.

Meredith, J.E., Jr., Fazeli, B., and Schwartz, M.A. (1993). The extracellular matrix as a cell survival factor. Mol. Biol. Cell *4*, 953–961.

Munger, J.S., Harpel, J.G., Giancotti, F.G., and Rifkin, D.B. (1998). Interactions between growth factors and integrins: latent forms of transforming growth factor-β are ligands for the integrin αvβ1. Mol. Biol. Cell *9*, 2627–2638.

Nunes, I., Gleizes, P.E., Metz, C.N., and Rifkin, D.B. (1997). Latent transforming growth factor-beta binding protein domains involved in activation and transglutaminase-dependent cross-linking of latent transforming growth factor-beta. J. Cell Biol. *136*, 1151–1163.

Prieto, A.L., Edelman, G.M., and Crossin, K.L. (1993). Multiple integrins mediate cell attachment to cytotactin/tenascin. Proc. Natl. Acad. Sci. USA *90*, 10154–10158.

Pytela, R., Pierschbacher, M.D., and Ruoslahti, E. (1985). Identification and isolation of a 140 kd cell surface glycoprotein with properties expected of a fibronectin receptor. Cell *40*, 191–198.

Sanford, L.P., Ormsby, I., Gittenberger-de Groot, A.C., Sariola, H., Friedman, R., Boivin, G.P., Cardell, E.L., and Doetschman, T. (1997). TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. Development *124*, 2659–2670.

Sato, Y., Tsuboi, R., Moses, H., and Rifkin, D.B. (1990). Characterization of the activation of latent TGF-beta by co-cultures of endothelial cells and pericytes or smooth muscle cells: a self-regulating system. J. Cell Biol. *111*, 757–763.

Schultz-Cherry, S., and Murphy-Ullrich, J.E. (1993). Thrombospondin causes activation of latent transforming growth factor-beta secreted by endothelial cells by a novel mechanism. J. Cell Biol. *122*, 923–932.

Schultz-Cherry, S., Lawler, J., and Murphy-Ullrich, J.E. (1994). The type 1 repeats of thrombospondin 1 activate latent transforming growth factor-beta. J. Biol. Chem. *269*, 26783–26788.

Schultz-Cherry, S., Chen, H., Mosher, D.F., Misenheimer, T.M., Krutzsch, H.C., Roberts, D.D., and Murphy-Ullrich, J.E. (1995). Regulation of transforming growth factor-beta activation by discrete sequences of thrombospondin 1. J. Biol. Chem. *270*, 7304–7310.

Sheppard, D., Cohen, D.S., Wang, A., and Busk, M. (1992). Transforming growth factor beta differentially regulates expression of integrin subunits in guinea pig airway epithelial cells. J. Biol. Chem. *267*, 17409–17414.

Shull, M.M., Ormsby, I., Kier, A.B., Pawlowski, S., Diebold, R.J., Yin, M., Allen, R., Sidman, C., Proetzel, G., Calvin D., et al. (1992). Targeted disruption of the mouse transforming growth factor-beta 1 gene results in multifocal inflammatory disease. Nature *359*, 693–699.

Sime, P.J., Xing, Z., Graham, F.L., Csaky K.G., and Gauldie, J. (1997). Adenovector-mediated gene transfer of active transforming growth factor-β1 induces prolonged severe fibrosis in rat lung. J. Clin. Invest. *100*, 768–776.

Taipale, J., Saharinen, J., Hedman K., and Keski-Oja, J. (1996). Latent transforming growth factor-beta 1 and its binding protein are components of extracellular matrix microfibrils. J. Histochem. Cytochem. *44*, 875–889.

Taipale, J., Saharinen, J., and Keski-Oja, J. (1998). Extracellular matrix-associated transforming growth factor-beta: role in cancer cell growth and invasion. Adv. Cancer. Res. *75*, 87–134.

Van Obberghen-Schilling, E., Roche, N.S., Flanders, K.C, Sporn, M.B., and Roberts, A.B. (1988). Transforming growth factor-beta1 positively regulates its own expression in normal and transformed cells. J. Biol. Chem. *263*, 7741–7746.

Wang, A., Yokosaki, Y., Ferrando, R., Balmes, J., and Sheppard, D. (1996). Differential regulation of airway epithelial integrins by growth factors. Am. J. Respir. Cell. Mol. Biol. *15*, 664–672.

Weinacker, A., Chen, A., Agrez, M., Cone, R.I., Nishimura, S., Wayner, E., Pytela, R., and Sheppard, D. (1994). Role of the integrin alphaVbeta6 in cell attachment to fibronectin. Heterologous expression of intact and secreted forms of the receptor. J. Biol. Chem. *269*, 6940–6948.

Werb, Z., Tremble, P.M., Behrendtsen, O., Crowley, E., and Damsky, C.H. (1989). Signal transduction through the fibronectin receptor induces collagenase and stromelysin gene expression. J. Cell. Biol. *109*, 877–889.

Woessner, J.F. (1961). The determination of hydroxyproline in tissue and protein samples containing small proportions of this amino acid. Arch. Biochem. Biophys. *93*, 440–447.

APPENDIX 3

OF
DECLARATION UNDER 37 C.F.R. § 1.132

DATED JANUARY 16, 2008

BY
STAFFAN JOHANSSON

Content:

Peer review of a manuscript submitted to the Journal of Biological Chemistry in 1993 (2 pages)

Response by the authors (4 pages)

Final decision by the editor (1 page)

MS# M3-3813
Forsberg and Johansson

This manuscript describes the preliminary and initial purification of another member of the integrin super family, which was named here alpha9beta1 (α9β1). The authors also characterize some of the peptide/protein binding properties of α9β1 and they use other rat tissues to attempt to localize its occurrence. While the authors may have purified a new protein, there are some problems associated with their observations and I have attempted to summarize my major criticisms below (I feel that the authors have done some interesting preliminary data on α9β1 yet I also feel that their results are too preliminary and the data is not strongly supported by quantitative observations):

1. The description of the protocol used was minimally described, for instance, what temperature was used throughout the homogenization?, what temperature and flow rate were employed for the column chromatography?, what matrix volumes were used for these preparation of insolubilized-ligands?, what were the amounts of proteins and peptides coupled to these matrices?, what sort of 'control matrix' was used to measure non-specific adsorption of these integrins past the WGA-Sepharose matrix? (this is especially important due to the lack of any apparent specificity of α9β1 to interact with immobilized RGD-containing peptides and the modified derivatives found in this work).

2. What is the status of the cysteine sulfhydryl group used in the peptide and the matrix? What it effectively deprotected following its synthesis to liberate the sulfhydryl moiety? Is it disulfide-linked to another peptide?

3. What percent acrylamide gel was used in these studies?

4. What is the specific activity of the ^{125}I-labeled integrin used in these studies? It would be irrpossible to reproduce this information described in the paper and to quantitate the binding data shown without such information.

5. The gel shown in Figure 2 has been cut-off very close to the Mr 90kDa marker, and it is not possible to assess what protein is visualized below this marker...inspection of Figure 1A reveals numerous lower molecular weight species in the silver stained gel...are these also present in Figure 2?

6. The sequencing data reported in the manuscript gives no data regarding the yield of each cycle sequenced for the alpha chain of your integrin. This is important information due to the routine difficulty of many labs in sequencing directly from blots as you have shown in your manuscript. Did you also sequence the beta chain since you had the purified protein on a blot? did it correspond to the exact sequence found for beta1?

7. Since you state that the conservation of integrin homology is normally >80% throughout various species, why didn't you analyze some human tissues or human cell/cell lines for the existence of α9β1? One would venture that it would be simpler and more informative than the procedure used here of analyzing other rat organs...this is especially important since you already have a polyclonal antibody to the amino-terminus of your integrin, and this would allow you to quantitate your observations, and not rely on the destruction of organs and potential low recovery of your and other integrins. It would also be much informative to perform specific tissue histochemical analyses to directly locate α9β1.

8. One could also speculate that your integrin alpha chain is not new, but a product of alternative splicing which is quite prevalent (it seems) in the integrins and extracellular matrix/cytoskeletal proteins....Is it possible that your alpha chain is just a variant of an existing alpha chain product? One could easily probe some liver libraries to detect the cDNA and obtain full-length alpha chain sequence, especially since you already have an existing specific polyclonal antibody.

9. The peptide specificity of α9β1 is intriguing, yet the preliminary observations need to be expanded to better understand these studies. The dilemma of why α9β1 would elute off an RGD-matrix, interact with immobilized laminin and then not be competed-off again from this surface by RGD-containing peptides is quite interesting yet the description is too qualitative. Regarding the studies of Figure 6, is the difference seen in binding due to a specific interaction of α9β1 with the various proteins or is it due to a difference in the actual amounts of ligands bound to the polystyrene surface? What would a BSA-control coated plate do in such a binding assay? It would most interesting to rigorously study the binding properties of α9β1 with various RGD-peptides, and highly defined fragments of laminin and fibronectin, these should all be part of the initial description of α9β1 for publication.

UPPSALA UNIVERSITET
Institutionen för medicinsk och fysiologisk kemi

Uppsala, October 14

Dr. Robert L. Hill
The Journal of Biological Chemistry
9650 Rockville Pike
Bethesda, MD 20 814
USA

Dear Dr. Hill,

Thank you for the rapid handling of our manuscript no M 3-3813 "Purification and characterization of integrin α9β1" by Forsberg et al. We have now revised the paper as suggested by the reviewer. We found his comments to be fair and helpful for improvement of the paper. In addition to further technical information, new experimental work have been added to the manuscript. Enclosed in this letter is a package including one copy of the previous version with xerox copies of the figures, and three copies of the revised version containing full sets of original pictures.

Our response to the points raised by the referee is listed below.

1. The temperature under which the purification was performed, flow rates during affinity chromatography, the amount of proteins and peptides coupled to Sepharose, and bed volume of columns have been included in "Materials and Methods" (p. 9).
In the paper, the integrin retained on GRGDSPC-Sepharose is shown not to bind to Sepharose conjugated with GRGDS, GRGDSP, the 105 kD fibronectin fragment, collagen, or laminin. Further, none of the integrins that we have studied bind to BSA-, IgG-, or thiol-Sepharose (not shown in the paper).

2. The sulfhydryl groups in the peptides used in this manuscript were all deprotected as described in the reference given in "Materials and Methods". To investigate the status of the sulfhydryl groups, two sets of experiments were performed, which are described in the manuskript on pages 7-8. The results demonstrate that the peptides are present as monomers on the gel matrix. Furthermore, DTT was applied to the GRGDSPC-Sepharose after the EDTA elution of integrin, and the material released by the reducing agent was analyzed by silver-staining after SDS-PAGE. No proteins could be detected in the DTT fractions. This result indicates that α9β1 binds directly to the peptide, rather than to a thiol-binding protein in the WGA-pool (p.15).

3. SDS-PAGE was performed in 7% acrylamide gels, which is now mentioned in "Material and Methods"(p. 8).

4. The specific activity of ^{125}I labeled α9β1 used in the solid phase receptor assay (~10 000 cpm/ng protein) has been added to the legend of figure 7 (p. 27).

Uppsala University
Department of
Medical and Physiological Chemistry

Postadress
Mailing address
Biomedicinska Centrum
Box 575
S-751 23 UPPSALA
Sweden

Gatuadress
Address
Husargatan 3

Telefon
Telephone
Nat 018 – 17 40 00
Int + 46 18 17 40 00

Cable
"Biomedicum"

Telex
76132 Biomed S

5. The small cut off picture was chosen since pictures of SDS-gels of our preparations of $\alpha 5\beta 1$ and $\alpha 1\beta 1$ have been published previously (Johansson et al. 1987, J. Biol. Chem. 262, 7819-7824; Forsberg et al. 1990, J. Biol. Chem. 265, 6376-6381). In the revised manuscript this picture has been replaced with a full-size photo of a new SDS-PAGE. In lanes 3 and 4 a protein of 70/75 kD (non-reduced/reduced) is seen in addition to $\alpha v\beta 3$. This protein is obtained in variable amounts and is derived from the blood that remains in the tissue after it has been cut into pieces and washed (mentioned on p. 13).

6. For all amino acid sequences shown in the paper, the repetitive yield was 92-95% for each reaction (p. 10).
We have now obtained partial amino acid sequences of the $\alpha 9$-associated β-unit ($\beta 1$), and of the αv-associated β-unit ($\beta 3$). These sequences are presented in fig. 4.

7. The peptide-antibody is working in immuno-precipitation and in western blots of rat $\alpha 9$, but unfortunately it is too weak for immuno-histochemistry. Also, the reactivity with human cells is poor. Therefore we have to await the generation of better reagents for the determination of the detailed location of $\alpha 9$. So far, we only know that it is a widespread integrin, which previously have been overlooked, and in some cases may have been taken as $\alpha 2$.

8. We have now obtained an amino acid sequence from an internal peptide of $\alpha 9$. This sequence is homologous to other α-subunits, and is probably located close to the middle of the protein in a region where the different α-subunits are well conserved (presented as fig. 3B). Since $\alpha 9$ has unique sequences at two distantly located sites, it is unlikely to represent a splice variant of the previously identified integrin subunits.
We are currently trying to obtain cDNA clones of $\alpha 9$ by screening of libraries and by PCR. This work is not always straight forward, and we have not been successful yet.

9. The coating efficiency of the proteins and fragments used in the solid phase assay have previously been found not to differ significantly except for collagens, which is adsorbed less efficiently at low concentrations (Timpl, Johansson, van Delden, Oberäumer & Hook 1983, J. Biol. Chem. 258, 8922-8927; Perris & Johansson 1987, J. Cell Biol. 105, 2511-2521; Perris, Paulsson & Bronner-Fraser 1989, Dev. Biol. 136, 222-238; Perris, Krotoski & Bronner-Fraser 1991, Development 113, 969-984).
Nevertheless, the dishes coated with collagen were good substrates for cell adhesion and binding of ^{125}I-labeled $\alpha 9\beta 1$.
BSA-coated wells do not bind the integrin $\alpha 9\beta 1$ and are used as a negative control (now mentioned in the legend to fig. 7).
In order to perform more detailed competition experiments of ligand-binding to the receptor, we tried to coat the wells with the RGDSPC peptide coupled to BSA or ovalbumin via glutaraldehyde. Unfortunately, this method did not allow for detection of $\alpha 9\beta 1$-binding to the peptide. However, we were able to locate regions in EHS-laminin and collagen I to which $\alpha 9\beta 1$ binds (fibronectin is not a ligand for this integrin, fig. 7A). Of the fragments CB3, CB7, and CB8, which comprise ~50% of the collagen $\alpha 1$(I) molecule, $\alpha 9\beta 1$ was found to bind only to the CB8 fragment. Two large

5. The small cut off picture was chosen since pictures of SDS-gels of our preparations of $\alpha 5\beta 1$ and $\alpha 1\beta 1$ have been published previously (Johansson et al. 1987, J. Biol. Chem. 262, 7819-7824; Forsberg et al. 1990, J. Biol. Chem. 265, 6376-6381). In the revised manuscript this picture has been replaced with a full-size photo of a new SDS-PAGE. In lanes 3 and 4 a protein of 70/75 kD (non-reduced/reduced) is seen in addition to $\alpha v\beta 3$. This protein is obtained in variable amounts and is derived from the blood that remains in the tissue after it has been cut into pieces and washed (mentioned on p. 13).

6. For all amino acid sequences shown in the paper, the repetitive yield was 92-95% for each reaction (p. 10).
We have now obtained partial amino acid sequences of the $\alpha 9$-associated β-unit ($\beta 1$), and of the αv-associated β-unit ($\beta 3$). These sequences are presented in fig. 4.

7. The peptide-antibody is working in immuno-precipitation and in western blots of rat $\alpha 9$, but unfortunately it is too weak for immuno-histochemistry. Also, the reactivity with human cells is poor. Therefore we have to await the generation of better reagents for the determination of the detailed location of $\alpha 9$. So far, we only know that it is a widespread integrin, which previously have been overlooked, and in some cases may have been taken as $\alpha 2$.

8. We have now obtained an amino acid sequence from an internal peptide of $\alpha 9$. This sequence is homologous to other α-subunits, and is probably located close to the middle of the protein in a region where the different α-subunits are well conserved (presented as fig. 3B). Since $\alpha 9$ has unique sequences at two distantly located sites, it is unlikely to represent a splice variant of the previously identified integrin subunits.
We are currently trying to obtain cDNA clones of $\alpha 9$ by screening of libraries and by PCR. This work is not always straight forward, and we have not been successful yet.

9. The coating efficiency of the proteins and fragments used in the solid phase assay have previously been found not to differ significantly except for collagens, which is adsorbed less efficiently at low concentrations (Timpl, Johansson, van Delden, Oberoumer & Hook 1983, J. Biol. Chem. 258, 8922-8927; Perris & Johansson 1987, J. Cell Biol. 105, 2511-2521; Perris, Paulsson & Bronner-Fraser 1989, Dev. Biol. 136, 222-238; Perris, Krotoski & Bronner-Fraser 1991, Development 113, 969-984).
Nevertheless, the dishes coated with collagen were good substrates for cell adhesion and binding of ^{125}I-labeled $\alpha 9\beta 1$.
BSA-coated wells do not bind the integrin $\alpha 9\beta 1$ and are used as a negative control (now mentioned in the legend to fig. 7).
In order to perform more detailed competition experiments of ligand-binding to the receptor, we tried to coat the wells with the RGDSPC peptide coupled to BSA or ovalbumin via glutaraldehyde. Unfortunately, this method did not allow for detection of $\alpha 9\beta 1$-binding to the peptide. However, we were able to locate regions in EHS-laminin and collagen I to which $\alpha 9\beta 1$ binds (fibronectin is not a ligand for this integrin, fig. 7A). Of the fragments CB3, CB7, and CB8, which comprise ~50% of the collagen $\alpha 1$(I) molecule, $\alpha 9\beta 1$ was found to bind only to the CB8 fragment. Two large

fragments of EHS-laminin, E8 and P1, were also tested in this assay. α9β1 bound specifically to E8 while no binding to P1 occurred. The three collagen fragments and the two laminin fragments were all good substrates for hepatocyte adhesion, further indicating that major differences in coating efficiency are not the reason for the different binding of ^{125}I-labeled α9β1. These data are presented in Figures 7B and 8. The binding of the integrin to EHS-laminin and to collagen I was specific in the sense that it was inhibited (60%) by a 50 fold excess of unlabeled α9β1. Further, the binding to collagen I was reduced by 70% if the protein was incubated at 50 C° for 10 minutes, and then coated at 37 C°. This is now mentioned in "Results" (p. 16).

We hope that these additions to the manuscript make it acceptable for publication in the Journal of Biological Chemistry. Please send editorial corresponence to the undersigned person, preferably by fax.

Yours sincerely

Erik Forsberg

The Journal of Biological Chemistry
Published by
American Society for Biochemistry and Molecular Biology

ROBERT L. HILL
Associate Editor

Department of Biochemistry
Duke University Medical Center
Nanaline H. Duke Building, Room 244
Research Drive
P.O. Box 3711
Durham, NC 27710
Tel: (919) 684-5164
Fax: (919) 684-5040

November 17, 1993

Dr. E. Forsberg
Dept. of Medical and Physiological Chemistry
Univ. of Uppsala
The Biomedical Center, Box 575
S-751 23 Uppsala, SWEDEN

Dear Dr. Forsberg,

Your manuscript with Drs. Ek, Engstrom and Johansson entitled, "Purification and Characterization of Integrin α9β1", has been reviewed by the reviewer who read the original version. I am sorry to inform you that he did not find the additional changes in the text sufficiently persuasive to change the recommendation to decline the manuscript. He did acknowledge that the additional description of the methodology and the new data improved the manuscript, yet also felt that more definitive studies characterizing this new integrin and its adhesion properties would be necessary. While most of the major methodology criticisms of the original critique were answered, it should also be mentioned that within the rapidly changing and emerging field of integrin biochemistry, it was felt that your studies did not permit a full enough comparison with other known integrins. Furthermore, the competition data with the peptides and protein fragments were interesting, yet it also requires further scientific development. The reviewers felt that you did substantially improve the manuscript, yet the original major concerns were not fully answered in your revised manuscript. In view of these negative opinions, I must decline your manuscript for publication in the Journal. I regret the need for a negative decision, but thank you for submitting the manuscript to the Journal.

I am not returning the manuscript unless you ask but please find enclosed the artwork.

Sincerely yours,

Robert L. Hill
Associate Editor

RLH/drm